12-15-2017

Genome Evolution in Najas and Hydrilla (Hydrocharitaceae).

Ursula King
University of Connecticut - Storrs, ursula.king@uconn.edu

12-15-2017

Genome evolution in Najas and Hydrilla (Hydrocharitaceae).

Ursula King

Genome Evolution in *Najas* and *Hydrilla* (Hydrocharitaceae).

Ursula Mary King, PhD

University of Connecticut, 2017

The aquatic monocot family Hydrocharitaceae, with 17 genera, displays considerable morphological and life history diversity. My research has focused on two cosmopolitan genera within this family, *Najas* and *Hydrilla*; and ranged from the species to family level. Chapter 1 characterized genetic diversity in the New World species, *Najas guadalupensis*. This widespread taxon currently is subdivided into four different subspecies in the Flora of North America. Evidence was provided here for extensive introgression into the *N. guadalupensis* genome from two North American congeners, *N. canadensis* and *N. flexilis*. Additionally, evidence was provided to suggest that current infraspecific taxonomic designations in *N. guadalupensis* are unwarranted.

The majority of angiosperm plastomes are conserved in size, gene content and order; however, a small number of lineages have been identified with aberrant plastomes. Chapter 2 focused on the evolution of the plastid-encoded polymerase genes (PEP) in *Najas*, and showed that two of these genes have unusual substitution patterns in this genus. In chapter 3, nine complete *Najas* plastomes (spanning both subgenera, *Najas* and *Caulinia*) were sequenced, assembled, and characterized; and a partial plastome assembly for the invasive aquatic *Hydrilla*

verticillata was also provided. Large scale movement of the inverted repeats in *Najas,* rearrangements in the large single copy region in *Hydrilla*, along with a mutual subset of highly divergent and missing genes, suggest that both *Najas* and *Hydrilla* possess atypical plastomes. In this chapter I also compared one hundred conserved and aberrant angiosperm plastomes, to determine whether there are any patterns evident in plastid genomes with aberrant evolution.

Molecular data to date have been unable to fully resolve relationships within Hydrocharitaceae. In chapter 4, plastid protein coding and ribosomal RNA (rRNA) genes for *Najas* and *Hydrilla* were incorporated with those from 14 other hydrocharit genera for a phylogenetic analysis. Apart from trees constructed with the full concatenated dataset, a number of separate analyses, based on different functional subsets of chloroplast genes (photosynthesis, plastid-encoded polymerase, ribosomal protein, and rRNA genes) were performed. These analyses provided evidence for conflicting phylogenetic signal between the different functional categories; with all genera within subfamily *Hydrilloideae* represented by very long branches with the RNA genes.

.

Genome Evolution in *Najas* and *Hydrilla* (Hydrocharitaceae).

Ursula Mary King

B.A., University of Dublin, Trinity College, Ireland, **2009**

M.A., University of Dublin, Trinity College, Ireland, **2012**

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2017

APPROVAL PAGE

Doctor of Philosophy Dissertation

Genome Evolution in *Najas* and *Hydrilla* (Hydrocharitaceae).

Presented by

Ursula Mary King, B.A., M.A.

Major Advisor

_____

Donald H. Les

Associate Advisor

_____

Cynthia S. Jones

Associate Advisor

_____

Yaowu Yuan

University of Connecticut

2017

Acknowledgments

I am deeply grateful to so many people for making this PhD possible, as I could never have achieved it on my own. I would especially like to thank my supervisor Donald Les, for introducing me to this wonderful group of plants, and for all his guidance and support along the way.

To my advisory committee, Elizabeth Jockusch, Cynthia Jones and Yaowu Yuan, I would also like to give a very special thank you. Their doors were always open for me, and they were so generous with their time and helpful advice.

I would like to thank everyone in the Ecology and Evolutionary Biology department for the many wonderful opportunities I have had here. I am so grateful to have had the opportunity to take so many wonderful courses in this department. To all the terrific EEB staff, Pat Anderson, Kathy Tebo, Anne St Onge and Lois Limberger; thank you for helping me navigate all the administrative necessities, with smiles and lots of laughter. Bob Capers and Jane O'Donnell in the Collections; Clinton Morse, Matt Opel and Adam Histen in the greenhouse; the laboratory technicians, Karen Lombard and John Korte; and Nicholas McIntosh; thank you for all your help in preparing for classes over the years.

A heartfelt thank you to my lab mates and friends Elena Peredo and Lori Benoit. It was just wonderful to spend a working day with both of you. I would also like to thank Jane Les, who did everything she possibly could for me when I first arrived in Connecticut, ensuring that I had everything that I needed, and that I got to see all the sights.

To Geert and Kathleen Goemans, and Beth and Josh Wade, thank you so much for welcoming me into your families and making me feel at home. For my wonderful friends, Rafael Medina,

Alfredo Manteca, Veronica Mantovani Bueno, Holly Brown, and Charlie Delavoi; and my dear housemates and friends, Juan Cabrera and Phuong Ha Nguyen; thank you for making this such a wonderful experience.

To Debbie Bryam, Robert Jellen, Jeremy Chartier and Judy Melichar, thank you for doing all that you possibly could to make me feel at home in Storrs. I would especially like to thank David Fried for always looking on the bright side, encouraging me, and for lots of laughter along the way. Mary Casey, Marie Coleman, and Mick Duffy, go raibh míle míle maith agaibh.

Finally, without the support and encouragement of my whole family, this would never have been possible. For my brothers Aidan and Martin, who have always looked after their "little sister"; and their families, Frances, Joan, Scott, Matthew, Colleen and Sean; my Dad, who in his quiet way is a rock of strength, wisdom and inspiration to all of us, and my Mam, who before she departed this world, gave all of us a love of plants, people and life. Finally, I want to thank my children, William, Emily and Thomas, who delight and inspire me with all they do. Thank you so much for the encouragement to take on this project. I am reminded every day how lucky I am to have all of you around me.

I would also like to acknowledge and thank my funding sources; including, the National Science Foundation (NSF DEB-0841658), the Irish Research Council for Science, Engineering and Technology (IRCSET), the Botanical Society of America, along with the University of Connecticut Graduate School, the Ronald Bamford Endowment Fund, and the Ecology and Evolutionary Biology Department.

I had assistance with parts of this project from Elena Peredo, Lori Benoit, Hamid Razifard and Cassandra Ball.

Is féidir linn!

Table of Contents

Chapter 1

Genetic diversity in *Najas guadalupensis* (Sprengel) Magnus (Hydrocharitaceae) in North America

Abstract

Aquatic vascular plants are characterized by extreme morphological reduction and phenotypic variation rendering it difficult to evaluate diversity in widely distributed aquatic species. Numerous attempts have been made to interpret the extensive morphological variation evident in *Najas guadalupensis* s.l. (Hydrocharitaceae), a New World species extending from Canada to Brazil. Currently, four subspecies are recognized in the Flora of North America. One of these, *N. guadalupensis* subsp. *muenscheri*, previously was found to be an allotetraploid hybrid derivative of *N. flexilis* (maternal) and *N. guadalupensis* that is synonymous taxonomically with *N. canadensis* Michx. This study focuses on the remaining infraspecific taxa recognized in North American *N. guadalupensis*. Considerable genetic variation and differences in flower and seed production were found within races of *N. guadalupensis* subsp. *guadalupensis*, along with widespread hybridization and introgression from the two annual species, *N. canadensis* and *N. flexilis*. Plants recognized as *Najas guadalupensis* subsp. *olivacea* also were found to be hybrid derivatives of *N. guadalupensis* (maternal) and *N. flexilis*, and that subspecies no longer is accepted here. Incongruence with respect to chloroplast and nuclear nrITS markers indicates that *N. guadalupensis* subsp. *floridana* also is of hybrid formation, at least in North America. In this case, *N. guadalupensis* subsp. *guadalupensis* is implicated as the maternal parent with an unknown (and previously unsampled) taxon as the paternal progenitor. Further sampling outside of North America will be necessary to resolve the parentage of this taxon with greater certainty.

It also is evident that nomenclatural priority was not observed properly in the original publication of var. *floridana* (on which subsp. *floridana* was based) with *N. flexilis* var. *curassavica* A.Br. (1864) being the correct basionym.

Introduction

Many groups of aquatic vascular plants are characterized by extreme morphological reduction along with extensive vegetative phenotypic variation in response to the aquatic environment (Sculthorpe 1967). Combined, these factors have hampered many systematic and taxonomic treatments, which were based solely on comparative morphology. Additionally aquatic plants are characterized by widespread distributions (de Candolle 1855; Sculthorpe 1967; Hutchinson 1975), rendering comprehensively evaluation of diversity across the entire range of individual species difficult. *Najas* L. (Hydrocharitaceae) is one such group. Comprised of about 30-40 diclinous (unisexual) species (Rendle 1899, Triest 1988), *Najas* is cosmopolitan in distribution and unusual in that it is one of only five freshwater plant genera (Les 1988) to complete reproduction entirely under water (hypohydrophily). Even amongst this hydrophilous group *Najas* is unusual in having such a high number of species, the other genera only having between one and six species (Les 1988). Additionally, most aquatic plants are perennial and readily propagate through clonal reproduction (Philbrick and Les 1996); however *Najas*, being predominantly annual, is believed to be one of the few exceptions to this rule (Hutchinson 1975).

Najas guadalupensis sensu lato

Najas guadalupensis (Sprengel) Magnus is one taxon within the genus, which is characterized by extensive morphological variation. Commonly known as Southern naiad, Southern water nymph or Guppy grass, *N. guadalupensis* s.l., has a widespread distribution extending from Canada to

Argentina (Clausen 1936, Haynes 1979, Lowden 1986) (Figure 1) and is believed to have originated in the Neotropics where the greatest extent of New World naiad speciation is thought to have occurred (Lowden 1986). Cottom (1938) demonstrated that *N. guadalupensis* is an important food source for waterfowl, which have been shown to be effective dispersal agents in other *Najas* species (Agami & Waisel 1986). Regarded as a nuisance weed in certain areas of North America (Anderson 1990, Hellquist 1997), *N. guadalupensis* is often associated with man-made irrigation ponds, canals and rice fields (Mason 1957, Stuckey 1971) and as its popular name "guppy grass" suggests, has long had an association with the aquarium and fish rearing trade (Műhlberg 1982, Kasselmann 2003).

Two subgenera historically have been recognized within *Najas* (Magnus 1870, Rendle 1899): *Najas* L. (*N. major* All. and *N. marina* L.) and *Caulinia* (Willd.) A. Braun (all remaining taxa). Magnus (1870) originally created two unranked groups (*Americanae* and *Euvaginatae*) within *Caulinia* and debate continues to what degree *Caulina* should be subdivided. Nonetheless, a number of authors recognized the cohesiveness of the *Americanae* group, treating it taxonomically as a section (Rendle 1899; Les et al. 2010; Ito et al. 2017). The focal taxon of this study, *N. guadalupensis*, is placed within the *Americanae* group along with *N. arguta* H.B.K., *N. candensis* Michx., *N. filifolia* R. R. Haynes [=*N. conferta* A. Br. *sensu* Lowden (1986)], *N. flexilis* (Willd.) Rost. & Schmidt., *N. microcarpa* K. Sch., and *N. wrightiana* A. Br. With the exception of *N. canadensis* (also in Eurasia), all of these taxa are restricted to the New World where a number of them occupy sympatric ranges. *Najas guadalupensis* co-occurs with *N. flexilis* and *N. canadensis*, principally above the Pleistocene glacial boundary in North America (Les et al. 2015) and with the remaining species in at least some portion of their American ranges (Lowden 1986, Urquiola Cruz 1988).

In summing up the Americanae Magnus (1870) wrote:

"Alle diese so mannigfach ausgebildeten Arten der Americanae werden, wie gesagt, durch die vielgestaltige N. microdon A.Br. verbunden, so dass jede weitere Theilung der Americanae unnatürlich erscheinen muss. N. microdon A. Br. bietet uns ein ausgezeichnetes Beispiel, wie ein Typus zugleich zu mehreren verschiedenen Typen Verwandtschaft haben kann, zu Typen, die gleichsam von ihm divergiren."; i.e.,

"All these highly varied species of *Americanae,* as we have said, are characterized by the diverse *N. microdon* A.Br. [= *N. guadalupensis*] so that any further division of the *Americanae* must appear unnatural. *N. microdon* A. Br. gives us an excellent example of how one type can also have kinship to several different types, to types which, as it were, diverge from it."

Chase (1947) attempted to correlate the effects of polyploidy with the extensive phenotypic variation in northern *N. guadalupensis* populations. Although he was unsuccessful in this endeavor, his study (Table 1), along with that of Davenport (1980) and a recent study by Ito et al. (2017), was invaluable in demonstrating that *N. guadalupensis* is represented by a diverse complex of diploids, tetraploids, hexaploids, heptaploids, octoploids and possibly enneaploids. Additionally, Chase's careful observations led him to propose hybridization in the formation of certain *Najas* taxa, a suggestion that has been supported in recent studies (Les et al. 2015).

In 1979, Haynes undertook a revision of the genus in North and Central America with Lowden (1986) revising the taxonomy of Neotropical *Najas* soon afterwards. These studies contributed much to understanding the overall distribution of *N. guadalupensis* s.l. in the New World, with both authors echoing the sentiments of previous authors such as Clausen (1936), who suggested that material of *N. guadalupensis* represented a 'decidedly heterogeneous aspect'. Prior to the North American revision (1979), two northern *Najas* species had been described, one from Minnesota, *N. olivacea* Rosendahl & Butters (1935), and the other from the Hudson river,

New York, *N. muenscheri* Clausen (1937). In his revision Haynes (1979) chose to recognize these two species as varieties of *N. guadalupensis*, along with another variety that he had earlier described from Florida, *N. guadalupensis* var. *floridana* Haynes & Wentz (1974). Lowden (1986), on the other hand, recognized just two infraspecific taxa in the Neotropics, both at the rank of form. One corresponded to Haynes' previously described variety from Florida (*N. guadalupensis* forma *floridana* (Haynes and Wentz) Lowden, comb. nov.) with everything else being designated as *N. guadalupensis* forma *guadalupensis*.

Table 1. Chromosome reports for *Najas guadalupensis* s.l. from Chase (1947), Davenport (1980) and Ito et al. (2017). The monoploid number for *Najas* is 6.

State	County	Lake	2n	Collector
N. guadalupensis* subsp. *guadalupensis				
Florida	Dade	35 miles west of Miami	48	*Davenport 960*
N. guadalupensis* subsp. *guadalupensis				
Alabama	Cherokee	Weiss Lake	24	*Davenport 457*
Alabama	Jackson	Guntersville Lake	24	*Davenport 473*
Alabama	Blount	Allgood village Creek	36	*Davenport 433*
Michigan	Mecosta	Horsehead Lake	48	*Chase* M-28
Michigan	Oakland	Lakeville Lake	54	*Chase* M-20
New York	Ulster	Mirror Lake	42	*Chase* NY-54, 80, 98, 99
New York	Columbia	Kinderhook Lake	42	*Chase* NY-49, 204, 210
New York	n/a	n/a	36	*Chase* NY-78
Texas	n/a	n/a	12	*Chase* T-4.
Japan	n/a	cultivated	ca. 48	*Ito* Y, 1142 & al.
N. guadalupensis* subsp. *muenscheri				
New York	Dutchess	Hudson river	24	*Chase* NY-58
N. guadalupensis* supsp. *olivacea				
Minnesota	n/a	Material in Cornell greenhouse	36	*Chase* Minn-1

As only a single infraspecific taxon is allowed in the Flora of North America, the rank of subspecies was chosen for *Najas*, necessitating that the initial varietal designations used by

Haynes in his revision be elevated to the rank of subspecies (Haynes & Hellquist 1996). Consequently, four infraspecific taxa of *N. guadalupensis* were recognized in the Flora of North America (Haynes 2000), which were delineated principally by seed, anther locule number, and leaf characters. *Najas guadalupensis* subsp. *guadalupensis* is the most widespread taxon, ranging from Canada to South America; *N. guadalupensis* subsp. *muenscheri* (R. T. Clausen) R. R. Haynes & Hellquist, was a Hudson River endemic in New York (Haynes 2000); *N. guadalupensis* subsp. *olivacea* (Rosendahl & Butters) R. R. Haynes & Hellquist was restricted to the northern and eastern United States and adjacent Canada along with *N. guadalupensis* subsp. *guadalupensis* (R. R. Haynes & Wentz) R. R. Haynes & Hellquist occurring further southward in Alabama, Florida, Georgia (Haynes 2000), Guatemala, Haiti, the Dominican Republic (Lowden 1976) and Cuba (Urquiola Cruz 1988).

N. guadalupensis* subsp. *guadalupensis

In addition to being the most widespread taxon, subsp. *guadalupensis* is characterized by extensive morphological variation marked by overlapping characters (Lowden 1986). Haynes and Wentz (1974) concluded that much of the variability in *N. guadalupensis* was due to phenotypic plasticity and was not genetically based. Lowden (1986), however, with reference to *N. guadalupensis* in the Neotropics, offered a somewhat more tempered opinion, concluding that further genetic studies were needed to determine natural limits within the group. Phenotypic plasticity in response to the aquatic environment has long been recognized in aquatic plants (Schuthorpe 1967; Hutchinson 1975); however, caution must be exercised in simply attributing extensive morphological variation in widespread taxa to environmental plasticity, without an assessment of how genetic diversity within a taxon is structured (Les & Philbrick 1993; Santamaría 2002).

Additionally, although subsp. *guadalupensis* has long been recognized as indigenous to the southern U.S. states, the native status of northern populations has been questioned in the past (Fernald 1908; Clausen 1936). Les and Merhoff (1999) reviewing the historical records of *N. guadalupensis* in New England suggested that it was probably native, but cautioned that it should be monitored as it was becoming locally aggressive in some parts of its range.

N. guadalupensis* subsp. *floridana

When the southern taxon subsp. *floridana* originally was described as a variety, Haynes and Wentz (1974) distinguished it from the typical variety, var. *guadalupensis* by its longer fruit (1.6-2.3 mm for var. *floridana* and less than 1.6 mm for var. *guadalupensis*), longer leaves (2-3.5 cm long in var. *floridana* and smaller in the typical variety) and by the number and prominence of teeth on the sides of the leaves (18-42 macroscopic teeth in var. *floridana* with "about 100 minute teeth" in var. *guadalupensis*).

Lowden (1986) clearly expressed a different opinion and chose to designate the two forms "merely on the visibility of teeth along the margin of leaf blades" stating that "no reliable correlation was found between this trait and other criteria, such as fruit length, number of teeth along leaf blades and geographic distribution". Unfortunately, while Lowden recorded a variation in seed length of 0.8-3.0 (commonly 1.0-2.5) mm, he provided no ranges for marginal teeth number or leaf dimensions.

Later, in the *Flora of North America*, Haynes (2000) used a combination of the following traits: leaf blade teeth evident to the unaided eye and anther locule number of one (subsp. *floridana*) versus leaf blade teeth invisible to the unaided eye and anther locule number of four (subsp. *guadalupensis*), to distinguish between these two taxa.

N. guadalupensis* subsp. *olivacea

In the closely related species *N. canadensis* and *N. flexilis*, the outer layer of the testa or seed coat becomes thick-walled and persists giving the seed a smooth polished appearance (Rendle 1899, Chase 1947); whereas, in *N. guadalupensis* subsp. *guadalupensis* (and subsp. *floridana*), this outer layer collapses and is partially lost forming distinctive rows of pits or areolae. Rosendahl and Butters (1935) established *Najas olivacea* based on the distinctive plants they found in Norway Lake, Minnesota (now extirpated). There they observed that *N. olivacea* differed from its congener *N. flexilis* in having non-lustrous, blunter, slightly falcate seeds with much larger, thicker-walled areolae, and from *N. guadalupensis* by its larger seeds and especially in the outer cells of the seed coat, which lacked the characteristic pitting of *N. guadalupensis*.

They also noted that no staminate flowers were observed at the time of their original collection in 1933 and that description of the species had to be deferred until male flowers were collected in the subsequent year. Later Chase (1947) reported a ploidy of $2n = 36$ for *N. olivacea*. Although he observed that its seed coat character suggested an admixture of *N. flexilis* genes, he concluded that *N. olivacea* represented a seed coat mutant of *N. guadalupensis* and that, except for the seed coat characters observed by Rosendahl and Butters, "no other anatomical or morphological feature ... could definitely distinguish this taxon from *N. guadalupensis"*.

Later when Haynes (1979) reduced *N. olivacea* to a variety of *N. guadalupensis* he concluded that var. *olivacea* represented populations of *N. guadalupensis* at the northern limit of its range. Surprisingly, neither his description nor key for the *Flora of North America* (Hayes 2000) mentioned the seed coat character, which arguably was the most distinguishing feature of subsp. *olivacea*. Instead, Haynes chose to distinguish this taxon from subsp. *guadalupensis* on

the basis of its leaf blade tooth number and stem width (20-40 teeth per side and a stem of 1 mm or more in diameter for subsp. *olivacea* vs. 50-100 teeth per side with stems of 0.8 mm or less for subsp. *guadalupensis*).

In the subsequent discussion, evidence is presented to suggest that subspecific status is taxonomically inappropriate for subsp. *muenscheri* or subsp. *olivacea*, and that proper nomenclatural priority was not observed in the naming of subsp. *floridana*. Consequently, these taxa will be referred to hereafter simply as '*muenscheri*', '*olivacea*' and '*floridana*'.

Hybridization

In the northern part of its range, *N. guadalupensis* overlaps the distributions of the annual species *N. canadensis* and *N. flexilis* (Haynes 2000, Les et al. 2015). Les et al. (2010) first documented hybridization in this group from three aggressive populations in Connecticut. Subsequently, they provided genetic and morphological evidence that *N. canadensis* was an allopolyploid derivative of *N. flexilis* and *N. guadalupensis*, that originated in sympatry (Les et al. 2015). As *N. canadensis* is abundant in Pleistocene records from Europe and both parental taxa are restricted to the New World, they hypothesized that it originated by means of an ancient hybridization event that occurred in North America, with later dispersal to Eurasia, a conclusion also congruent with DNA sequence data. That study also revealed that the putative Hudson River endemic, *N. guadalupensis* subsp. *muenscheri* ('*muenscheri'*) was nested within the *N. canadensis* clade with respect to both genetic and morphological evidence. Consequently, this taxon will not be considered any further in the present study.

Reproductive potential

Najas guadalupensis is the most widespread *Najas* species in North America (Haynes 2000) and yet we lack a great deal of basic information about life history strategies employed by this taxon. Northern populations are often recorded as sterile or semi-sterile (Fernald 1923; Clausen 1936; Chase 1947). Yet Rendle (1899) recorded fruits in nodal clusters of two to five fruit in his study based on a limited number of southern U.S., Caribbean, Central and South American accessions.

Stuckey et al. (1978) provided valuable information on the ability of *N. guadalupensis* to overwinter at northern latitudes. Towards the end of the severe winter of 1976-1977 in Ohio, when air temperatures remained below zero for the month of January and below 10°C for the duration of February, they recovered *N. guadalupensis* from under 30 cm of ice and demonstrated by subsequently growing these plants in aquaria that they had remained physiologically active. That demonstration provided good evidence that some populations of *N. guadalupensis* were perennial in the northern part of its range. A search of the Southeast Regional Network of Expertise and Collections portal (http://sernecportal.org) reveals a number of southern populations collected during the winter months, which also suggests that some populations at southern latitudes also are perennial. Yet, *N. guadalupensis* sometimes is reported as an annual species (e.g., DiTomaso & Healy 2003). Given the evidence above, at this point we can assume that *N. guadalupensis* in North America is at least facultatively perennial; however, given the widely varying reports on fertility between northern and southern populations, this still leaves a large amount of uncertainty about the degree to which various populations are channeling resources towards sexual versus vegetative reproduction. *Najas guadalupensis* is unusual within the genus in evolving an ability to perennate and having a better understanding of some basic aspects of life history strategies in this taxon should further our understanding of

evolution in this group. Given that most aquatic plants are perennial, with clonal propagation being more common than sexual reproduction (Hutchinson 1975, Les 1988), one obvious question is, are all *Najas* species really exceptions to this rule?

Study aims

Highly reduced morphological traits and high infraspecific variation (often attributed to phenotypic plasticity) seriously hinders taxonomic resolution in submersed aquatic plants, especially those with widespread distributions. With the availability of molecular markers, the levels of genetic diversity in these taxa may be unraveled and in many instances may uncover high levels of cryptic variation (e.g., Zhu et al. 2015, Neiva et al. 2017). The present study aims to characterize genetic variation in *N. guadalupensis* s.l. in North America, to determine how that variation is structured and to test whether molecular data support current infraspecific taxonomic designations. The employment of genetic data already has enabled detection of cryptic hybridization and speciation in *Najas* (Les et al. 2010; 2015, Ruegg et al. 2015); and now allows for earlier hypotheses of hybridization events, based on morphological intermediacy or chromosome variation, to be tested. Specifically:

- How is genetic diversity structured in *N. guadalupensis*?
- Are current infraspecific taxonomic designations warranted?
- Is hybridization with *N. canadensis* or *N. flexilis* implicated in the origin of *N. guadalupensis* subsp. *olivacea*?
- To what extent does the current range of *N. guadalupensis* in North America reflect natural versus human-mediated processes?

Characterization of reproductive variation in *N. guadalupensis* is an important step in determining whether certain populations are principally behaving as low fertility clonal perennials, as opposed to sexual perennials, or indeed as obligately sexual annuals. Annual species, like *N. canadensis* and *N. flexilis*, are expected to have a higher seed output than the perennial species (e.g., van Kleunen 2007), as successful fruiting is essential for their survival. Because a large herbarium collection of these three species has been compiled as part of the systematic study of the genus in North American, this resource provides an ideal opportunity to compare reproductive output across all three closely related species. Including the annual species will provide a base line for comparison with *N. guadalupensis*, to assess how reproductive potential varies between the perennial and two annual species. Additionally, as the allotetraploid *N. canadensis* has been taxonomically separated from *N. flexilis* only recently, a comparison of reproductive output between these two cryptic species will be facilitated. Extensive sampling across the North American range of *N. guadalupensis* should help to disclose whether all northern populations are characterized by low reproductive output (as reported for some populations) or whether reproductive output is more likely to be heterogeneous across all *N. guadalupensis* populations. If considerable variation is observed, then a comparison of genetic and location data may help to evaluate whether variation in fertility results primarily from temperature or latitudinal effects or whether other (i.e., genetic) factors might be involved.

Materials and Methods

Specimen collection

Field work, targeting the North American distribution of *Najas guadalupensis* (Haynes 2000) was conducted between 2007 and 2013. Collections times were centered on the months of July

and August to maximize the potential of finding flowering/fruiting material, as this is the best way to distinguish between *Najas* species (Braun 1864). Plant specimens were preserved in NaCl/CTAB solution (Rogstad 1992) with corresponding dried voucher specimens deposited in the CONN herbarium (Appendix A). Additional specimens were obtained from collaborators and processed similarly. Field specimens were georeferenced on-site using a GPSmap76SC unit (Garmin International, Olathe, KS, USA) or georeferenced manually using locality information provided by the collectors. Herbarium material was sampled with permission (CDA, CONN, MOAR, UC, WI) including an isolectotype of *N. guadalupensis* subsp. *olivacea* (MICH 1485091).

DNA isolation, sequencing and analysis

Total genomic DNA from NaCl/CTAB preserved material was extracted using the CTAB method of Doyle and Doyle (1987). A modified CTAB method for herbarium specimens was used, following Les et al. (2013). Sequence data for three loci were obtained for 197 accessions following Les et al. (2010). These included two chloroplast regions: *rbcL* and *trnK*/*matK* and the nuclear internal transcribed spacer (nrITS) region. These loci had provided good infraspecific resolution in previous studies of *Najas* (Les et al. 2010, Les et al. 2015). Additionally, to estimate whether further substructure could be determined, the chloroplast *trnL-F* region was sequenced for a subset of 46 accessions. The *trnL-F* region was amplified with the universal c and f primers of Taberlet et al. (1991) in a 12.5µl total reaction volume using 0.15 mM each dNTP (Promega, Madison, WI, USA), 0.4 µM each primer, 1x Titanium Taq® reaction buffer with 0.065 µl Titanium Taq® polymerase (Clontech, Mountain View, CA, USA) and 20 ng of template DNA. Thermal cycling conditions consisted of 3 min initial denaturation at 94°C, then

30 cycles of denaturation at 94°C (1 min), annealing at 55°C (1 min), and extension at 72°C (3 min), with a final extension at 72°C for 5 min.

nrITS universal primers

Most *N. guadalupensis* accessions produced polymorphic nrITS sequences with the universal ITS5 and ITS4 primers (White et al. 1990), as evidenced by double peaks on electropherograms. The high prevalence of polymorphic sequences precluded subcloning of this region from all individuals; therefore, as new ribotype combinations were encountered, individuals were resequenced and PCR products were subcloned following Les et al. (2010). Eighteen accessions in total were subcloned with between eight and 10 sequences obtained for each of the subcloned individuals. To minimize the influence of any PCR recombination that may have occurred during the cloning process, only ribotypes that were present in more than two individual clone PCR's were considered.

nrITS repeat-specific primers

Examination of the seed morphology of the lectotype (MIN 1001844) and an isolectotype of *N. guadalupensis* subsp. *olivacea* (MICH 1485091) indicated possible admixture from *N. flexilis* or *N. canadensis*. To determine whether this was the case, two sets of repeat-specific primers were designed (Nguad-ITS and Nflex-ITS) which would amplify a shorter fragment of the nrITS region to discriminate between variation in *N. flexilis*, *N. canadensis* and *N. guadalupensis* and potentially have a higher success of amplification in herbarium material. The program Primer3 (Untergasser et al. 2012) as implemented in Geneious V6.1 (http://www.geneious.com, Kearse et al. 2012) was used to design these repeat-specific primers. PCR reactions were as above for the *trnL-F* region, with the addition of 1.25 ul dimethyl sulfoxide (DMSO) (Sigma-Aldrich, St.

Louis, MO, USA) to minimize secondary structure problems. Thermal cycling conditions for the Nguad-ITS primers comprised: 2 min initial denaturation at 94°C, then 30 cycles of denaturation at 94°C (40 s), annealing at 62°C (50 s) and extension at 72°C (50 s), with a final extension at 72°C for 3 min. A lower annealing temperature of 60°C was used for the Nflex-ITS amplicons.

As sequencing of products generated by these two primer sets indicated hybridization between *N. flexilis* and *N. guadalupensis* in the type material for subsp. *olivacea*, and further sampling demonstrated hybridization between some northern populations of *N. guadalupensis* and either *N. flexilis* or *N. canadensis*, all *N. guadalupensis* accessions subsequently were scored with these primers using gel electrophoresis to explore whether hybridization was more extensive within *N. guadalupensis*. As a control measure, PCR's were carried out with both primer sets, along with positive and double negative controls (water and DNA) in all PCR reactions. PCR products were screened for bands after electrophoresis in 1% agarose gels, stained with SybrGreenI, with a subset of 79 accessions selected for sequencing of these shorter amplicons. In addition to amplifying *N. flexilis*/*N. canadensis* amplicons, as sampling was expanded, it was discovered that the Nflex-ITS repeat-specific primers also amplified the subsp. *floridana* repeat type, even though mismatches were present in the priming sites. However, as the initial expectation was that southern populations of *N. guadalupensis*, outside of the current distributional ranges of these two species, would give a null amplification, amplification of the subsp. *floridana* allele was deemed acceptable at that point.

For all amplified regions, fragments were purified using an equal volume of PCR product and diluted (1:4) ExoSAP-IT® enzyme mixture (Affymetrix, Inc., Santa Clara, California). Cycle sequencing for all regions consisted of 2.0 µl purified amplification product with 0.5 ul BigDye® V1.1 (Applied Biosystems, Foster City, CA, USA), 2 ul of 5x ABI buffer and 0.35 uM

of sequencing primer in a 7 µl reaction. Final products were cleaned with sephadex colums, using 600 ul of a 6.5g SephadexTM G-50 mix (GE Healthcare Bio-sciences AB, Uppsala, Sweden) and sequenced with an ABI PRISM 3100 genetic analyzer (Applied Biosystems, Foster City, CA, USA).

In all instances where new haplotypes or ribotypes were uncovered, accessions were re-sequenced to eliminate the possibility of any PCR artifact or sequencing error. Additionally a number of accessions were re-extracted and sequenced directly from herbarium specimens to confirm results and to verify that initial sampling of CTAB preserved material comprised only one individual.

Data editing and alignments

Sequence chromatograms were checked manually using CodonCode Aligner V1.2.1 (CodonCode Corporation, Dedham, Massachusetts) and sequences were assembled into contigs. For the universal ITS region, where the majority of individuals produced polymorphic sequences (and subcloning of all of these accessions would have been prohibitively expensive), a combined approach was used. Here, information derived from a number of individuals, which were subcloned for this region, and others, which were amplified with the N-guad ITS repeat-specific primers, facilitated the reading of chromatograms manually and assisted with the scoring of individual ribotypes encountered in polymorphic sequences. However, because the repeat-specific primers amplified a shorter fragment of the nrITS region, the total number of nrITS variants in *N. guadalupensis* might be underestimated. All sequences were checked for the three conserved Viridiplantae 5.8s motifs (M1: CGATGAAGAACGyAGC, M2:

GAATTGCAGAAwyC and M3: TTTGAAyGCA) following Harpke & Peterson (2008) for the identification of possible pseudogenes.

Alignments

Using only the non-redundant chloroplast haplotype and nrITS sequences from the universal primers, alignments were conducted using MUSCLE v3.8.31 (Edgar 2004), as implemented in Geneious, with the default settings, and visually inspected for translation of coding regions and indel alignment wherever relevant. Insertions and deletions (indels) were scored for the alignments using the simple indel coding scheme of Simmons and Ochoterena (2000) as implemented in the program SeqState V1.4.1 (Müller 2005). A 138 nt region of the *trnL-F* locus was removed in the final alignment due to difficulty in confidently aligning indels over this region.

Phylogenetic analysis

Sequence data from a previous study (Les et al. 2015) representing sequence diversity in *N. canadensis* (n = 9) and *N. flexilis* (n = 9) and for additional outgroup taxa, *N. filifolia* [= *N. conferta*] (n = 1) and *N. wrightiana* (n = 1) were incorporated in the analyses. All initial analyses were conducted using these four outgroups; however, as *N. filifolia* comprised a long branch and the same trees were obtained with or without inclusion of this taxon, *N. filifolia* was removed in the final analyses.

PartitionFinder v1.1.1 (Lanfear et al. 2012) was used to determine optimal partitioning schemes, with respective best fit substitution models. A number of *a priori* partitioning schemes were assessed as follows: a) single partition, b) by gene, with separate partitions for tRNA, spacer and intergenic regions, where appropriate, c) by codon position, with separate partitions

for tRNA, spacer and intergenic regions and d) by gene and by codon position with separate partitions for tRNA, spacer and intergenic regions, where appropriate. Model search was restricted to those models implemented in MrBayes and the relative fit of each partitioning scheme was assessed using the Bayesian information criterion (BIC) score (Schwarz 1978; Sullivan & Joyce 2005).

Individual gene regions were analyzed separately using maximum likelihood implemented in GARLI v2.0 (Zwickl 2006) and Bayesian inference as implemented in MrBayes v3.2.6 (Ronquist et al. 2012). All analyses were run on the University of Connecticut's BBC Bioinformatics Facility Cluster. For maximum likelihood analyses, 30 independent search replicates were run to investigate whether tree space was being thoroughly sampled. To estimate branch support, 500 bootstrap replicates were performed from single random starting trees. Bootstrap support was summarized using SumTrees v.3.3.1 in the DendroPy v.4.0.3 python package (Sukumaran & Holder 2010). For Bayesian inference, two independent runs of 10 million generations, each with four simultaneous chains, and a discard burn-in of 25%, were run (see Appendix B for partition parameters). Tracer v1.5 (Rambaut and Drummond 2007) was used to examine convergence of the parameter values and chain stationarity.

Preliminary analyses indicated that the chloroplast loci (*rbcL*, *matK* and *trnL-F*) were congruent and these regions were concatenated for the final analysis. However, as initial sequencing had indicated that a number of different nrITS ribotypes were associated with individual chloroplast haplotypes, both within each chloroplast haplotype group and within certain individuals, the nrITS datasets from both the universal nrITS primers and the repeat-specific primers were analyzed separately.

For the non-redundant ribotypes obtained using the universal nrITS primers, trees were constructed with maximum likelihood and Bayesian inference as described above. However, sequences obtained using the Nflex-ITS repeat-specific primers indicated that these primers were potentially amplifying pseudogenized nrITS copies in some taxa, as sequence diversity in the conserved 5.8s region was greater than might be expected within a given species or closely related group of taxa. To assess relative divergence among sequences associated with these repeat-specific primers and to compare with that already encountered in *N. canadensis*, *N. flexilis* and *N. guadalupensis* with the universal primers, a NeighborNet equal angle network from uncorrected p-distances was constructed using SplitsTree4 v4.14.4 (Huson and Bryant 2006). Included in this analysis were sequences of individual ribotypes observed in *N. flexilis*, *N. canadensis* and *N. wrightiana* (Les et al. 2015), and all sequence variants for *N. guadalupensis* obtained with the universal ITS primers (and NguadITS repeat-specific primers) in this study, along with all sequences obtained for *N. guadalupensis* using the Nflex-ITS repeat-specific primers. Recombination in this region was assessed using the pairwise homoplasy index (PHI statistic, Bruen et al. 2006), also implemented in SplitsTree4.

Reproductive potential

Using dried herbarium specimens, 177 accessions of *N. guadalupensis* s.l. were evaluated for reproductive potential. Additionally, accessions of *N. canadensis* (n = 78) and *N. flexilis* (n = 49) which had been collected and sequenced either for this study or our previous study (Les et al. 2015) were also evaluated. *Najas* specimens are often intertwined on herbarium sheets making it difficult to confidently follow individual stems from upper nodes to lower nodes. Therefore, to standardize counts in relation to the variability of herbarium material, the top 12 nodes of a single mature stem per plant were scored for the presence of male and female flowers, seed

number, and the presence of adventitious roots. Additionally, the entire specimen was also scored for the presence/absence of reproductive material, as this may not have been observed on the stems counted; and whether seed, if present, was hollow or otherwise appeared to be inviable. *Najas* plants can produce adventitious roots at the nodes. An increase in the production of adventitious roots potentially could indicate more of an investment in vegetative reproduction and so the presence of adventitious roots at nodes was also noted. Finally, collection dates, and whether the population was indicated as aggressive by the collector, were recorded. We assume that our large sample size should give a good indication of reproductive potential; however, we acknowledge that a more rigorous study might have incorporated averages of a number of stems from each plant.

To gain finer resolution among *N. guadalupensis* genotypes, we hoped to evaluate all sequenced accessions; however, a number of specimens could not be assessed due to either insufficient material, concealment of flowers/fruit by herbarium mounting glue, or the unavailability of herbarium loans which had previously been returned. The minute nature and simplicity of the flowers precluded discriminating between male and female flowers from available online digital images of these specimens.

Seed and leaves from a representative number of individuals were imaged using a Leica MZ16 dissecting microscope connected to a JVC KY-F75U digital camera with additional images taken using an AmScope MD900 dissecting microscope. Data were visualized using ggplot2 (Wickham 2009), as implemented in Rstudio (RStudio Team 2016).

At the time of writing, I became aware that priority might not have been observed in naming of *N. guadalupensis* subsp. *floridana*, and that the distribution of subsp. *olivacea* was

apparently far more restricted than previously reported (Haynes 2000). At that point, time only allowed online searches of digitized specimens. I was aided in this part of my work by the invaluable service provided by the Consortia of Midwest Herbaria (http://midwestherbaria.org), the Northern Great Plains Herbaria (http://ngpherbaria.org), the Integrated Digitized Biocollections portal (https://www.idigbio.org) and the Southeast Regional network of Expertise and Collections (http://sernecportal.org). Likewise, the service provided by the Biodiversity Heritage Library (https://www.biodiversitylibrary.org) was invaluable in accessing older literature.

Results

Specimen collection

Two hundred and thirty two samples of *Najas guadalupensis* s.l. representing 199 different North American localities were obtained for this study, with two additional samples from Costa Rica and Honduras procured from colleagues. North American collections fell mainly within the distributional range reported by Haynes (2000) with the exception of the states of DE, ME, MT, NV, OR, RI, TN, UT and VA (although surveyed). Additionally, new records are reported from NM and WY.

Model selection

The combined chloroplast alignment length of 38 non-redundant haplotypes consisted of 3579 nts and 45 indels and represented 19 haplotypes from *N. guadalupensis* s.l. (197 accessions) and 19 associated with outgroup taxa. The optimal partitioning scheme for this alignment reflected gene codon position, with separate partitions for the tRNAs, introns and intergenic regions.

The alignment length of 25 non-redundant sequences of the nrITS region (universal primers) was 748 nts with 12 indels. Fourteen of these ribotypes represented *N. guadalupensis* and 11 were associated with the outgroups. Optimal partitioning selected for this region was by gene, with the internal transcribed spacers as a separate partition. Table 2 summarizes these schemes and the associated diversity information for *N. guadalupensis* and outgroups. Alignment region and sequencing primer information is included in Appendix B.

Table 2. Optimal partitioning schemes and chosen models and the number of analyzed sites in the combined chloroplast haplotype alignment and the nrITS dataset, along with related genetic diversity information, including uninformative variable (VS) and parsimony informative (PI) sites.

Partition	Model	No. sites	VS	PI
Chloroplast				
rbcL 1 + *trnK*/*matK*_1	JC+I	524	1	1
rbcL 2 + *trnK*/*matK*_2	F81+I+G	522	1	4
rbcL 3 + *trnK*/*matK*_3	HKY+I	522	7	6
tRNA-K + tRNA-L introns	HKY+I	1453	28	24
tRNA-L + tRNA-F	JC	65	0	0
trnL-F intergenic spacer	F81	493	12	9
Indels	Standard/1 rate	45	22	23
Total		**3579**	**71**	**67**
	nrITS			
18S+5.8S	JC	192	2	0
ITS1+ITS2	HKY	556	46	42
Indels	Standard/1 rate	12	4	8
Total		**748**	**52**	**50**

Chloroplast variation in *Najas guadalupensis sensu lato*

The more conservative *rbcL* region held the least variation with only three haplotypes recovered and was unable to resolve major clades within *N. guadalupensis* (Appendices C and D). One haplotype was represented by northern and eastern populations; another had a widespread North American distribution and a third haplotype corresponded to a number of southern populations

but was also found in populations from Kansas, Nebraska, and South Dakota. Our sample from Honduras and accessions of *'floridana'* both contained this latter *rbcL* haplotype.

The more variable chloroplast *trnK*/*matK* region consisted of 16 different variants in total, and provided some further geographic resolution. This marker differentiated subsp. *olivacea* as a northern variant of the more widespread *rbcL* haplotype and '*floridana'* as a variant of the more southern *rbcL* haplotype. Variation in the *trnL-F* region was congruent with that found in *trnK*/*matK* and added little additional resolution with the exception of three variants found in '*floridana'* and two in '*olivacea'*.

Combined chloroplast haplotypes were designated cG1–16, with the minor variants associated within '*floridana'* and '*olivacea'* (based on the *trnL-F* region) given the additional letter designations cG7 a,b,c and cG4 a,b respectively.

Phylogenetic analysis

Phylogenetic trees obtained by maximum likelihood (ML) and Bayesian inference (BI) were topologically congruent for the three chloroplast gene regions and most clades were similarly supported in both analyses (Figure 2). All three species, *N. canadensis, N. flexilis* and *N. guadalulpensis* resolve as strongly supported clades. Within *N. guadalupensis* s.l, the clade representing accessions from Honduras (cG1), Nebraska and South Dakota (cG2), California and Texas (cG3), and Kansas (cG4) also resolves as a strongly supported clade (BS: 95, PP: 1.0); but the cG5 haplotype is poorly supported as a separate lineage. This cG5 haplotype is comprised of two accessions from Texas, noteworthy for possessing numerous seeds (with up to three seeds per node in many instances). The remaining clade (cG6–16) represents the majority of sampled

North American *N. guadalupensis* accessions. This clade also included accessions with '*olivacea'* and '*floridana*' morphology.

Haplotypes representing all individuals from Florida with the '*floridana'* phenotype (cG7a,b,c) along with two individuals from Florida with intermediate leaf phenotypes between subsp. *guadalupensis* and '*floridana'* (cG6 and cG8) resolved as a distinct lineage (BS: 79, PP: 1.0). Sequencing of the chloroplast *trnL-F* region distinguished three minor variants within the '*floridana'* haplotype (cG7 a,b,c), based on single indels, but no overall support exists for any groupings within this clade. I was unable to obtain this region for all accessions in this clade (or the '*olivacea'* clade), so minor variants are indicated here but only the conserved designation cG7 is discussed.

The three remaining *N. guadalupensis* lineages resolved as a polytomy. Of these three lineages, one group (cG9–cG12) represented the majority of accessions within the widespread *rbcL* haplotype. The cG9 haplotype represents a single accession from Missouri, and of the three other haplotype lineages, cG10 represents the most widespread clade, occurring south of the Great Lakes region, with a widespread distribution on both sides of the Mississippi, as far south as Louisiana to the east and extending to California in the western part of its range. A single disjunct population located in Teton County, Wyoming is notable both in its discontinuous distribution within this haplotype, and also as it is outside of the previously reported range for *N. guadalupensis* (Haynes, 2000). The remaining two haplotypes (cG11 and cG12) overlapped with the more western part of the cG10 range, with cG12 having a discontinuous distribution between the states of Iowa, Nebraska and Kansas and those of New Mexico, Arizona and California; and cG11 represented by two accessions, one from Kansas and the other from Texas.

A second well-supported group represents individuals within the widespread *rbcL* haplotype, which were further discriminated by the more variable *trnK*/*matK* and *trnL-F* regions. The type material for *N. olivacea* fell within a clade representing haplotypes at the more northerly limit of this widespread group (cG14). All of our accessions from Minnesota and Wisconsin that corresponded to this phenotype also possessed this variant. One accession possessed a seven bp insertion in the *trnL-F* region (cG14b), differentiating it from the other '*olivacea'* haplotype; however, no morphological differences were evident. Also included in this cG14 clade were four accessions lacking the '*olivacea*' phenotype. All of these accessions were from Pennsylvania which is south of the previously reported range '*olivacea*' and were highly similar to each other in morphology, with elongated stem internodes and short leaves. However these samples lacked the stouter stems of '*olivacea*'. A lineage representing samples from California and Mississippi (cG13) resolved as sister to this '*olivacea*' haplotype clade with good support (BS: 95, PP: 1.0).

Haplotypes in the final clade resolved within *N. guadalupensis* (cG15–cG16), shared the northern/eastern *rbcL* haplotype and represent the largest number of individuals in our sampling. The *trnK*/*matK* and *trnL-F* regions resolved these two groups into a broad northern and eastern lineage (cG16), associated principally with the Great Lakes region, but also found in Washington state, southern Indiana, and down along the Atlantic states as far as South Carolina. A single anomalous individual from South Carolina with '*floridana*' morphology also possessed this haplotype. The second lineage (cG15) was represented by individuals found solely in Atlantic coastal states in our sampling.

Genetic variation in *Najas guadalupensis sensu lato* with universal ITS5 and ITS4 primers

Analysis of nrITS variation in *N. guadalupensis* indicated 14 different ribotypes, designated nrG1–nrG14. However, in total 61% of individuals (142/232) had polymorphic sequences, indicative of additive sequence variation. Trees obtained from both maximum likelihood and Bayesian inferences had similar support values and were topologically congruent, but many nodes had low to moderate support in both analyses. As several different nrITS ribotypes were associated with individual chloroplast haplotypes in *N. guadalupensis* s.l., the non-redundant nrITS phylogeny is presented with a heat map corresponding to the related haplotypes (Figure 3).

In contrast to the chloroplast analysis, both maximum likelihood and Bayesian inference resolved the ribotype associated with '*floridana*' (nrG1) as a distinct clade, separate to the one containing all remaining *N. guadalupensis* accessions (and the lineages representing *N. canadensis* and *N. flexilis*). All individuals associated with the '*floridana*' leaf phenotype possessed this distinct ribotype, including the accession from South Carolina (Najaflor001) possessing the anomalous northern chloroplast haplotype (cG16). This nrG1 ribotype differed from all other *N. guadalupensis* ribotypes by eleven substitutions and one indel (Appendix E). However, while all individuals with the '*floridana*' leaf phenotype possessed the nrG1 ribotype, it also occurred in five accessions from Alabama (1), Florida (2), South Carolina (1) and Wyoming (1) not corresponding to that morphology. Although the two accessions from Florida resolved within the '*floridana*' (cG6 and cG8) haplotype clade as mentioned, the other accessions from Alabama and Wyoming possessed the more widespread cG10 haplotype variant, with the South Carolina accession possessing the northern haplotype (cG16), similar to our other accession from South Carolina, which did exhibit '*floridana*' morphology.

The remaining *N. guadalupensis* ribotypes (nrG2–nrG14) are well separated as a monophyletic group (BS: 100, PP: 1). A well supported lineage, (nrG2–nrG6) is represented primarily by southern accessions; however, at the limits of their distribution, these ribotypes are found extending as far north as Butte County, California to the west, and West Virginia to the east. Within this clade, the nrG2 ribotype resolves as sister to the rest of the clade. This ribotype was found only in our two accessions from Texas (exclusively associated with the cG5 haplotype and notable for having two to three seeds per node). Relationships within the rest of this clade have only moderate support. The ribotypes nrG3 and nrG4 possessed wide overlapping geographic distributions and were mostly associated with the widespread chloroplast haplotypes (cG9-cG12), with the following exceptions: both alleles co-occurred in a single individual from North Carolina, with the northern chloroplast cG15 haplotype (no seed was evident) and the three accessions from California associated with the cG13 haplotype, which represented the more southerly accessions of the '*olivacea*' haplotype clade. Two of these three accessions possessed seeds. Additionally, one or the other of these two ribotypes co-occurred with the nrG1 ribotype in five of the '*floridana*' accessions and none of these accessions possessed seeds. The nrG5 ribotype was the only allele found in a single individual from southern Texas, which also possessed fruit. The last ribotype in this clade (nrG6) was again the only ribotype present in our accession from Honduras. Notably, it also was the only allele present in another herbarium accession from Costa Rica. Unfortunately (as these were our only two samples from Central America) it was not possible to obtain any of the chloroplast regions from the Costa Rican sample. The Honduran sample was only a tiny fragment and no seed was observed; however, the Costa Rican sample was one of the few other *N. guadalupensis* accessions that possessed

multiple seeds per node. All of the Texan samples and the Honduran sample within this ribotype clade fell within the southern chloroplast haplotype clades (cG1–cG4 and CG5).

Support for relationships within the third major *N. guadalupensis* clade (representing the majority of our samples) is very low (BS: 57, PP: 0.74). Within this clade, the ribotypes nrG7 and nrG8 grouped together. These ribotypes have a widespread distribution south of the Great Lakes region, with the majority associated with the widespread chloroplast haplotype clade (cG9–cG12). However, these alleles co-occurred in most individuals with ribotypes from another group in this clade (nrG9–nrG11). No accession associated with these two ribotypes resolved in the northern chloroplast haplotype clade (cG15–cG16). However, they did co-occur with the nrG1 allele in two accessions from the '*floridana*' clade (cG6–cG8), one accession with the '*floridana*' leaf phenotype, and the other lacking it. Neither of these accessions bore fruit. Additionally, this ribotype occurred in an accession from Texas in the cG3 haplotype clade without fruit.

Of the remaining two lineages, the nrG9–nrG11 grade is comprised of ribotypes that have a similar widespread distribution as the previous clade, but extend a little farther north (west of the Mississippi River as far as Brookings County, South Dakota and to the east as far north as Indian County, Pennsylvania) but again occurring primarily south of the Great Lakes region. Once more, the majority of accessions in this clade were associated with the widespread chloroplast haplotypes (cG9–cG12) and the cG13 clade; however, seven accessions from Indiana (n = 2), Ohio (n = 1) Pennsylvania (n = 1) and North (n = 1) and South Carolina (n = 2) possessed the northern/eastern haplotype cG16. Additionally, these ribotypes were associated with all of the northern accessions from the cG1–cG4 haplotype. This assemblage represented individuals from Nebraska (n = 2) and South Dakota (n = 3), with the cG2 haplotype along with

a single accession collected at Atchison State fishing lake in Kansas (cG4) in which both the nr9 and nr10 alleles co-occurred. No fruit was evident on this accession; however, four out of the five accessions from Nebraska and South Dakota possessed numerous seeds, which is noteworthy given that they were collected from quite northern localities. Three of the four accessions from the cG3 haplotype clade also possessed the nrG9 ribotype. These accessions were from Texas (n = 2) and California (n = 1).

The final ribotype lineage (nrG12–G14) was chiefly associated with the northern/eastern haplotypes (cG15 and cG16) with the exception of two individuals collected at LSU Aquaculture Research Station, Louisiana and Lake Murray, Oklahoma which possessed the widespread cG10 haplotype. Additionally, these ribotypes were also found in the haplotype clade represented by '*olivacea*' (cG14), with '*olivacea*' accessions all having the nrG14 ribotype and the four accessions from Pennsylvania, lacking the '*olivacea*' phenotype all possessing the nrG12 ribotype.

Along with amplifying *N. guadalpensis* alleles, an additional fifteen *N. guadalupensis* individuals also amplified either *N. canadensis* or *N. flexilis* alleles using the universal primers (Table 3). All of these individuals were from the cG15–cG16 haplotypes. Les et al. (2010) previously reported three hybrid individuals from aggressive populations in Connecticut. In this analysis, these individuals also fell within these clades. Additionally, four accessions originally collected as *N. guadalupensis* amplified an *N. canadensis* haplotype but possessed both *N. guadalupensis* and *N. canadensis* alleles.

Table 3. Eight North American states where *Najas guadalupensis* accessions possessing *N. canadensis* or *N. flexilis* ribotypes were found with the universal nrITS primers. Alleles amplified are indicated along with associated haplotype information.

Location	Haplotype	Ribotype a	Ribotype b
Connecticut (2)	cG16	nrG12	*N. flexilis*
Ohio (1)	cG16	nrG12	*N. flexilis*
Indiana (1)	cG16	nrG12	*N. flexilis*
Indiana (1)	cG16	nrG12	*N. flexilis*
Connecticut (1)	cG16	nrG12	*N. canadensis*
Michigan (1)	cG16	nrG12	*N. canadensis*
Pennsylvania (1)	cG16	nrG12	*N. canadensis*
Connecticut (3)	cG15	nrG12	*N. canadensis*
Connecticut (4)	*N. canadensis*	nrG12 + nrG13	*N. canadensis*

An additional 35 N. guadalupensis accessions were sequenced for the nrITS region and gel scored for the Nflex and Nguad-ITS repeat primers; however, chloroplast regions were not obtained for these accessions, or in a few cases, only the less variable *rbcL* region was sequenced because of either insufficient or degraded DNA from herbarium specimens. Included in this group were two additional '*olivacea*'accessions. However, overall patterns for the resulting ribotypes were the same for these accessions (Appendix C).

Analysis of ITS repeat-specific primers

a) Nflex-ITS

Unexpectedly, the Nflex repeat-specific primers also positively amplified regions in the majority of *N. guadalupensis* accessions. In addition to amplifying the *N. flexilis* allele, these primers also amplified the *N. canadensis* and '*floridana*' (nrG1) alleles, if present.

Direct sequencing of products amplified with the Nflex-ITS repeat specific primers in 79 *N. guadalupensis* s.l. accessions resulted in an alignment length of 438 nt. Network analysis resulted in six major clusters (Figure 4 [A] and Appendix F). These clusters corresponded to

alleles associated with *N. canadensis* (nrC), *N. flexilis* (nrF) and sequences which were polymorphic at the positions differentiating *N. canadensis* and *N. flexilis* (nrC + nrF). Additionally, one group of sequences were similar to *N. flexilis* (nrF-pseudo) but formed their own distinct cluster. These sequences are interpreted here as older copies of the *N. flexilis* repeat type which have become pseudogenized. Another cluster represented alleles associated with '*floridana*' (nrG1). Finally a highly divergent group (nrUn) represented a loose cluster of sequences that were most similar to *N. canadensis*. These sequences potentially could represent pseudogenized copies of *N. canadensis* alleles or partially concerted ribotypes. Alternatively, they may represent divergent, previously unsampled sequences of another *Najas* taxon. The closest blast hits on GenBank to these sequences are to *N. canadensis*. No statistically significant evidence of recombination (p=0.317) among these ribotypes was detected using the PHI statistic (Bruen et al., 2006), and with the exception of nrF-pseudo, all contained the conserved 5.8s motifs of Harpke & Peterson (2008).

All of the *N. guadalupensis* accessions with *N. flexilis* alleles were collected from within the known distributional range of the latter taxon (Les et al. 2015) and were represented by individuals with the northern chloroplast haplotypes (cG14 and cG16). Likewise, the accessions with ribotypes interpreted as being pseudogenized *N. flexilis* alleles also occurred in northern populations and all fell within the cG16 haplotype.

Accessions with the *N. canadensis* amplicon were associated with a number of different haplotype clades and had a more widespread distribution, as follows: the northern clades: cG14 [PA(3)], cG15[NC, CT(5), NJ], cG16 [IN, OH(2), CT], the widespread haplotype clades: cG10 [IA, MO,AZ(2)],cG12 [IA, NM], and the southern haplotype clades cG2 [NE(1)], cG3 [TX(2)], cG5 [TX(2)]. This result was somewhat unexpected, particularly because Arizona, New Mexico

and Texas are far south of the current distributional range for *N. canadensis* (Les et al. 2015) (Figure 4 [B]). Additionally, all of the samples from both the widespread haplotypes and the southern haplotypes with these alleles had seed, and in some cases, these alleles were present in the accessions with many seed.

All individuals with '*olivacea*' morphology (cG14) had alleles that clustered with the *N. flexilis* allele, whereas the *N. canadensis* allele was amplified from those individuals from Pennsylvania lacking this phenotype. A further two samples from Wisconsin and Minnesota with '*olivacea*' morphology also had the nrG14 ribotype and amplified an *N. flexilis* allele with the Nflex-ITS repeat primer. However, because these samples were from herbarium material and did not amplify any of the chloroplast regions, their haplotype could not be determined.

The nrG1 allele was the only allele amplified from accessions from the '*floridana*' clade (cG7–cG8) with these primers. However, this allele was also amplified from two individuals lacking '*floridana*' morphology (cG10 [WY], cG6 [FL]), a result not detected using the universal primers. Both contained the same polymorphic site in the 5.8s region, indicating the presence of two nrG1 alleles. Interestingly, both of these accessions were collected from extremely hot water springs.

These primers also amplified a divergent group of sequences (nrUn), which clustered most closely with *N. canadensis* in the analysis. These alleles had a relatively widespread distribution and associated with the southern/widespread chloroplast haplotypes (cG3, cG10, cG11, cG12, cG13) with the exception of a single individual in southern Indiana, which possessed the northern cG16 haplotype (but with widespread nrG10 and northern nrG12

ribotypes). No evidence of these nrUn alleles was found in either the '*olivacea*' (cG14) or '*floridana*' (cG7) haplotype clades.

b) Nguad-ITS

In all cases where the Nflex repeat primer was sequenced for an individual, the Nguad repeat primer was also sequenced. These sequences verified the initial ribotype scoring from the universal ITS primers and additionally, in a few cases, these primers amplified an additional *N. guadalupensis* allele that was not identified using the universal primers. However, none of these products was surprising, in light of the overall trends.

During the course of testing the Nguad repeat-specific primers, sixteen *N. canadensis* and *N. flexilis* accessions were screened. As an allopolyploid derivative of *N. flexilis* and *N. guadalupensis*, there was a possibility that *N. canadensis* might amplify a *N. guadalupensis* repeat type; however, no amplification was evident in this small sample.

Analysis of reproductive potential

Analyses of counts of seeds, male and female flowers and adventitious roots from the first 12 nodes of a single shoot revealed certain patterns with respect to *N. guadalupensis* s.l. and the two annual species *N. canadensis* and *N. flexilis*. The percentage of specimens with seed in both *N. canadensis* (69%) and *N. flexilis* (65%) was similar with the majority of individuals in both taxa possessing fruit, as opposed to *N. guadalupensis* where only 23% of individuals had seed. A similar pattern was also observed with respect to female flowers (Figure 5). The mean number of seeds per individual across the 12 nodes was 2.41 [±2.35 SD] for *N. canadensis*, 1.61 [±1.50 SD] for *N. flexilis* and 0.52 [±1.31 SD] for *N. guadalupensis*, with the mean number of female flowers being 1.12 [±1.39 SD], 1.06 [±1.45 SD], and 0.29 [±0.81 SD] for each taxon

respectively. However, the distribution of individual counts per specimen for both seeds and female flowers varied widely among the three species. Whilst the majority of *N. guadalupensis* individuals had no seed or female flowers, notably certain individuals had higher counts equaling those of *N. canadensis* and *N. flexilis*. The distributions within *N. canadensis* and *N. flexilis* were much wider, with fewer individuals lacking seed altogether and most possessing female flowers. In both *N. canadensis* and *N. flexilis* most seed and female flowers were distributed amongst the upper nodes (Figure 6) in comparison to *N. guadalupensis* where no particular pattern was observed.

The pattern with respect to male flowers was somewhat different. In this case, 36% of *N. guadalupensis* individuals possessed male flowers in contrast to *N. canadensis* (9%) and *N. flexilis* (20%). Additionally, *N. guadalupensis* was observed to have male flowers across all 12 nodes, with zero percentage of individuals having male flowers after node five in *N. canadensis* and after node three in *N. flexilis*. The mean number of male flowers for both *N. canadensis* and *N. flexilis* was 0.39 [±0.84 SD] and 0.14 [±0.55 SD] respectively. Whilst the distribution of counts per individual was much narrower in the latter two taxa, a much larger distribution was again evident within *N. guadalupensis* [mean 1.2 ±1.98 SD] where a number of individuals had high male flower counts, which exceeded those in both *N. canadensis* and *N. flexilis*.

All three taxa showed a similar propensity for nodal production of adventitious roots, with *N. canadensis* having a marginally higher percentage of specimens with this attribute (38%), followed by *N. guadalupensis* (29%) and *N. flexilis* (24%). While a few individuals of both *N. canadensis* and *N. guadalupensis* were observed to begin nodal production of roots at slightly earlier stages than *N. flexilis*, all three taxa showed adventitious root production around

nodes five to six. However, the highest percentage of individuals with roots at any given node was only 15%, indicating that overall production was quite low in these upper nodes.

The distribution of seed and flowers across *N. guadalupensis* haplotypes is given in Figure 7. For consistency, all haplotypes are included even though sample size is very low for some groups. Nevertheless, some trends and observations are worth noting especially for the haplotypes with large sample sizes.

Within the northern cG16 haplotype (n=59), only 2 individuals (3%) possessed seed. These represented our sample from South Carolina with the '*floridana*' ribotype (nrG1). In this case the seed was of questionable viability; however, seed on the other specimen from Ohio appeared to be viable. Female flower production also was low (5%) in that clade; however, in contrast 34% of accessions had male flowers, all of which were singular at a node, but often distributed along four or five nodes.

The second largest sample (n = 55) represented the geographically widespread haplotype, cG10. This group represented a heterogeneous mix with regards to the number of individuals with seed (38%), female flowers (20%), and male flowers (53%); and 12 specimens lacked both seed and flowers. A wide variety of seed sizes was also observed within this haplotype, with seed lengths varying from 1.1 mm to 2.2 mm (Appendix G). None of these results correlated with nrITS ribotypes or geography. While four individuals in this haplotype had many seed, most individuals only had a few, with no more than a single seed per node observed on any individual specimen.

Our sample size was low (n = 8) for the haplotype clade representing individuals with '*olivacea*' morphology (cG14) and low fertility overall was observed. Two samples,

corresponding to individuals with '*olivacea*' morphology (nrG14 + nrF) possessed seed. While seed appeared plump and viable, only two seeds in total were observed across the 12 nodes counted in these individuals. No male flowers were observed on these specimens. Regarding the four specimens from Pennsylvania (nrG12 + nrC), none possessed fruit and only two female flowers of questionable viability were observed; however, many male flowers were present on two of these individuals. Similarly, within the '*floridana*' haplotype clade, cG7 (n=9), fertility also was very low.

As mentioned, sample size from several haplotypes was very low and little can be inferred at this stage. Yet, having observed the range of variation within *N. guadalupensis*, two haplotypes (cG2 and cG5) are worth noting in terms of the quantity of seed observed, the number of seeds per node, their geographic locations and their genetic distance from the rest of the *N. guadalupensis* clade. The cG5 haplotype was represented by two specimens from Texas which were markedly different from other accessions by their copious, long seeds (>2mm), with many nodes possessing two or even three seeds (Appendix G). The other notable haplotype (cG2) represented individuals from Colorado, Nebraska and South Dakota, also with numerous seeds. This result is of interest given the northern locality of these specimens.

No seed was observed in a number of haplotypes (cG1, G4, cG6, cG8, cG9, cG11); however, sample numbers in all of these groups are low and little can be inferred from this observation.

Leaf and seed images

To illustrate the range of variability across *N. guadalupensis*, sample photographs of at least a single seed per specimen are provided in Appendix G (and Figure 4 (D)). All seed observed had

pitted seed coats, typical of subsp. *guadalupensis* and '*floridana*', with the exception of individuals within the '*olivacea*' clade (cG14 + nrF).

Leaf images of all individuals possessing the nrG1 ('*floridana*') ribotype are also provided along with a representative sample across *N. guadalupensis* accessions (Appendix H).

Discussion

Submersed aquatic plants with highly reduced morphologies represent a broad phylogenetic group (Les 1988). Elucidation of genetic diversity in this group is interesting both from a theoretical aspect in tracing evolutionary histories and processes, but also has important practical applications for conservation and the identification of potentially invasive non-indigenous genotypes (e.g., Moody & Les 2007, Les et al. 2013). This study has identified a number of genetic races in *N. guadalupensis* in North America with geographical structuring and raises the question whether reproductive output and life history strategies vary in these races. Recent studies have demonstrated hybridization between *N. guadalupensis* and the closely related annual species *N. canadensis* and *N. flexilis* (Les et al. 2010; Les et al. 2015). Indeed some of the many synonyms associated with *N. guadalupensis* (e.g., *N. flexilis* var. *curassavica* A. Br., *N. flexilis* var. *fusiformis* Chapman, *N. flexilis* var. *guadalupensis* (Sprengel) A. Br. reflect the difficulty that earlier taxonomists experienced in separating these species based on morphology. This study provides further evidence that the evolutionary histories of these three species have been intertwined through successive events of hybridization and admixture.

Najas guadalupensis subsp. *guadalupensis*

Chloroplast variation

Haplotype analysis determined three major well supported clades within subsp. *guadalupensis* with a degree of geographical patterning in North America, likely to be attributable to a number of factors including historical glaciation events, various dispersal vectors and potential secondary radiations of more divergent genotypes from Central American or Caribbean populations.

Najas pollen preserves poorly (Birks 2006) and few accessible records of *N. guadalupensis* macrofossils for glaciated areas in North America exist; however, evidence provided by Les et al. (2015), along with the discovery of interglacial macrofossils (min. 50,000 ybp) of both *N. guadalupensis* and *N. canadensis* at Ithaca, New York (Karrow et al. 2009) suggest that the predominantly southern *N. guadalupensis* (Clausen 1936, Lowden 1986) has had a long Pleistocene history of co-existence with *N. canadensis* and *N. flexilis* in North America, as far north as the Great Lakes region.

Fossil pollen profiles have provided evidence for a number of North American glacial refugia for aquatic plants during the Last Glacial Maximum, and demonstrate the rapid post-glacial recolonization of previously glaciated areas, as soon as seasonally ice-free habitat became available (Vesper & Stuckey 1977, Dieffenbacher-Kral & Jacobson 2001, Swada et al. 2003). Such rapid northward migrations of aquatic plants, between wetlands and against the flow of major rivers, likely were mediated by waterfowl for which aquatic plants act as a primary source of food and shelter (e.g., Amezaga et al. 2002, Soons et al. 2008). Both endozoochory (internal transport) and ectozoochory (external transport) are recognized as being important processes for the dispersal of these plants (reviewed in Figuerola & Green 2002), with Soons et al. (2008) estimating that mallard ducks can effectively disperse wetland plant seeds in relatively large

numbers, by up to 3000 km during migration. Aquatic plant refugia corresponding to Beringia, the Pacific Northwest, Atlantic coastal states and the Mississippi Embayment regions have been proposed (Dieffenbacher-Kral & Jacobson 2001, Swada et al. 2003). In eastern North America, the latter two refugia are believed to result from the major barrier to east-west aquatic plant migrations imposed by the Appalachian Mountains (Vesper & Stuckey 1977, Sawada et al. 2003).

Results here suggest that separate pools of subsp. *guadalupensis* populations survived the Last Glacial Maximum to the south in the Mississippi Embayment region (cG9–cG14 haplotypes and nrG7–nrG11 ribotypes) and to the east along Atlantic coastal states (cG15–cG16 haplotypes and nr12–nr14 ribotypes). It is proposed that Mississippi Embayment populations migrated northward postglacially along the Mississippi Valley radiating east and west along various tributaries, while a subset of these populations migrated west along the Rio Grande and into California. The isolated occurrence in Wyoming of the cG10 haplotype is likely to be a more recent introduction (discussed later). On the other-hand the cG15–cG16 haplotypes conceivably were isolated to the east of the Appalachian range, extending northwards in the early Holocene.

Which routes aquatic plants have taken to postglacially recolonize the Great Lakes region has been of considerable interest (Vesper & Stuckey 1977, Les et al. 2013). It has been suggested that source populations for this region could come from a) refuge populations in the Mississippi Embayment, with migrations northward along the Mississippi valley, radiating east and west at the confluence of the Ohio and Mississippi rivers (Vesper & Stuckey 1977), b) Atlantic coastal populations, which migrated northward to New England and westward from there (Vesper & Stuckey 1977, Sawada et al. 2003, Les et al. 2013) or c) later range extensions into the Great Lakes Region of relictual Pacific Northwest populations (Sawada et al. 2003). Evidence here

suggests that Great Lakes populations of *N. guadalupensis* were sourced from Atlantic coastal populations rather than Mississippi Embayment or Pacific Northwest populations, mirroring previous results for another *Najas* species, *N. gracillima* A. Br. (Les et al. 2013).

No evidence is found in our sampling for any divergent haplotypes which might implicate a Pacific Northwestern refugium. Samples from Washington State all shared the cG16 haplotype, and given that this haplotype clusters with the cG15 haplotype, which is found only on the east coast, these Pacific populations likely were sourced from Great Lakes populations. East-west North American disjunctions are common in aquatic plants (Les 1986), however, the remote nature of this haplotype is difficult to explain given contemporary north-south bird flyways. Nevertheless even modern flyways are apt to change temporally (Buhnerkempe et al. 2016), and certainly bird migration patterns during successive Pleistocene glacial-interglacial cycles must have changed considerably (Buehler et al. 2005). As the shrinking Laurentide and Cordilleran ice sheets broke apart, approximately 12,000 – 13,000 ybp (Sawada et al. 2003), more nesting habitats were provided to the west rather than the east, and late Pleistocene range expansions attributed to this are implicated in current indirect bird migration routes (Buehler et al. 2006). This factor may explain why today some birds with overwintering grounds in southern Atlantic coastal states take longer routes around the Great Lakes to breeding grounds in Eastern Canada. However, while this provides a plausible explanation for colonization of the Great Lakes region from eastern populations of *N. guadalupensis*, the Pacific Northwest populations are still difficult to explain. The extent of postglacial water bodies in the landscape and the effect of temperature oscillations during the Holocene (Viau et al. 2006) are both factors to consider, as indeed are rare long-distance dispersal events, attributable to vagrant birds (Cain et al. 200, Veit 2000). Unfortunately, although surveyed, no *N. guadalupensis* populations were located in the

states of Montana and Oregon which could help to give a more complete picture. Alternatively, of course, the possibility of a human-mediated introduction of this haplotype cannot be ruled out.

Given the large genetic divergence of the cG1–cG4 and cG5 haplotypes from the rest of the subsp. *guadalupensis* clade, it is conceivable that these lineages originate from secondary radiations of *N. guadalupensis* from Central America; or alternatively, these clades may represent separate isolated glacial populations within North America. All of the aquatic plants presented in the analysis by Swada et al. (2003) had a presence in Texas, at the mouth of the Rio Grande at 21,000 ybp, and so another North American refugium is plausible. These clades also represented isolated populations in Nebraska, South Dakota (cG2) and Kansas (cG4) and while greater sampling may have uncovered further southern populations associated with these haplotypes, these populations may represent isolated disjunctions. Certainly, these fertile Midwestern populations are of great interest, as are populations in Texas from the cG5 clade given their genetic divergence and high fertility.

Reproductive output appears to vary with haplotype but the extent to which this is associated with both ecological factors and genetic factors will require further study. Certainly, given that the cG15–cG16 haplotypes are found in southern Atlantic coastal states without fruit and yet populations from Nebraska and South Dakota corresponding to the cG2 haplotype have a high seed output, this appears to suggest that temperature alone may not be the sole factor contributing to low fecundity, as Fernald (1923) suggested. These populations presumably experience similar cold winter temperatures to northern populations from the cG16 clade.

nrITS Variation – Universal primers

The high occurrence of *N. guadalupensis* individuals exhibiting polymorphic sequences with the universal ITS primers was somewhat unexpected. Most individuals, with the exception of the cG1–cG5 haplotypes, had polymorphic nrITS copies throughout the species North American range. This result was in contrast to our previous results for the two annual species, *N. canadensis* and *N. flexilis*, where polymorphic sequences were encountered infrequently and only were associated with hybridization between two of these three species (Les et al. 2015). Lower ribotype diversity also was associated with these annuals with only a single ribotype found in the conterminous United States, and a second ribotype associated with relictual populations in Alaska. Similarly, just six ribotypes were associated with *N. canadensis* throughout its North American range, with two ribotypes recovered in Eurasian populations (from Ireland to Russia). It is difficult to explain such homogenization across wide geographic distances in these species, yet Rüegg et al. (2016) also observed complete nrITS homogeneity, and no diversity, within the dioecious annuals *N. major* and *N. marina* (sampled throughout a wide geographical region in Eurasia). Moreover, these two species are known to differ through chromosomal arrangements (Viinikka 2009); yet natural hybrid individuals also were recovered in the analysis of Rüegg et al. (2016).

Ribosomal repeats can undergo a number of different fates when two divergent genomes unite, from the maintenance of both repeat types, to various degrees of recombination resulting in chimeric repeats or complete concerted evolution to one repeat type (which may be unidirectional in favor of either of the parental contributors). Moreover, none of these processes are mutually exclusive which makes it difficult to draw conclusions on the tempo and direction of nrITS evolution (Malinska et al. 2010). These attributes may often confound phylogenetic

inference. However, where both repeat types are retained they can be highly informative with respect to historical relationships (reviewed in Álvarez &Wendel 2003).

A number of hypotheses may be invoked for this observed nrITS heterogeneity in *N. guadalulpensis*. One explanation might be that concerted evolution and homogenization of the nrITS region might occur at a slower rate, reflecting longer generation times in a perennial species (Soria-Hernanz et al. 2008), yet results across different plant groups have not been consistent with this hypothesis (reviewed in Gaut et al. 2011). Another factor might be the prevalence of agamic reproduction within a species; with slower rates of gene conversion frequently associated with asexual species (Campbell et al. 1997, Felliner & Rosselló 2007). Alternatively, if polyploidy is at play, the occurrence of several ribosomal DNA loci located on separate homeologous chromosomes may retard homogenization of repeats (Felliner & Rosselló 2007). Evidence exists that all three processes are likely to occur, to varying degrees, within *N. guadalupensis*.

Results here provide quantitative evidence to suggest that northern/eastern haplotypes (cG15–cG16) of *N. guadalupensis* are predominantly perennial and propagating clonally, with limited fertility. Potentially viable seed was only observed on two specimens from these clades; however, a high prevalence of male flowers was also observed. Pollen viability tests will be required to determine gene flow potential through male function in these populations. Nevertheless, the lack of fruiting in this large sample suggests predominantly clonal propagation, supporting previous observations of certain northern populations (Fernald 1923, Rosendahl 1935, Clausen 1936, Les et al. 2010). Additionally, potential turion (overwintering bud) production was recently discovered in a population of subsp. *guadalupensis* from a Connecticut Lake

(discussed below), providing additional evidence that these populations are perennating as predominantly clonal populations.

While no overall difference in adventitious root production was observed between *N. guadalupesis* and the two annual species (Figure 5), which might further indicate increased investment towards vegetative propagation, this may have been due to the limited number of plant nodes surveyed here.

In contrast are the highly fertile populations in Texas (cG5) and the Midwest (cG2). Plant life cycles, while often proffered as discrete conditions (Fox 1990), are often labile within a species (e.g. Barbier et al. 1989, Van Kleunen 2007) and it would be interesting to determine whether the Midwestern populations, in particular, are annual or perennial given their northern location. On the other hand, variable fertility in perennial plants might not be unexpected (Friedman & Rubin 2015); however, highly reduced fertility, along with observed nrITS heterozygosity across many subsp. *guadalupensis* haplotypes indicates that other factors, such as polyploidy may be responsible for fertility variation in *N. guadalupensis* (Soltis 2000).

Nflex repeat-specific primers

Undoubtedly, evidence exists for polyploidy playing a major role in the evolution of *N. guadalupensis*. Chase (1947) reported a range of chromosomal counts in populations from Michigan and New York (Table 1). These populations presumably fall within the cG16 haplotype here, in which all accessions had high heterozygosity with the universal nrITS primers (nrG12–nrG14), and showed extensive admixture from *N. canadensis* and *N. flexilis* with the repeat-specific primers. Polyploidy in plants may cause wide variation in sexual fertility (Eckhart 2003); but perenniality and the ability to propagate vegetatively could help buffer any negative

effects associated with increased ploidy (Grant 1971, Les & Philbrick 1993, Ramsey and Schemske 1998, Mallet 2007).

Although only a few chromosome counts exist for southern populations of subsp. *guadalupensis*, individuals from three populations in Alabama have been reported as tetraploids and hexaploids (Davenport 1980). In our analysis, populations from this state were within the widespread haplotype, where high variation in fertility and high nrITS heterozygosity also was recorded. Additionally, in the few accessions that failed to amplify a second ribotype, non-amplification due to PCR bias or drift, or complete locus loss, cannot be ruled out (Malinska et al. 2010).

On the other hand, the only diploids reported to date for subsp. *guadalupensis* have been from Texas (Chase 1947) and it is of interest that Chase's account of diploids in Texas correlate with populations here showing high fertility. Unfortunately, Chase failed to report a specific location for the Texan populations; and in this study, a total of four haplotypes were recorded from Texas (cG3, cG5, cG10, cG11), so little may be inferred. Additionally, while the highly fertile cG5 accessions only amplified a single repeat type with the universal nrITS primers, both accessions showed evidence of admixture from both *N. canadensis* and *N. flexilis* with the Nflex repeat-specific primers.

Within subsp. *guadalupensis*, it is impossible, from these data, to determine how much of this nrITS heterogeneity results from intraspecific hybridization accompanied by somatic doubling, or is related to autopolyploidy (polyploids associated with conspecific parents), with later divergence of one repeat type. However, it appears that at least some of the variation results from interracial hybridization in contact zones. It is also hard to determine the degree to which

allopolyploidy (interspecific hybridization and chromosome doubling) or autopolyploidy might be implicated in this range of chromosomal variation within *N. guadalupensis*. While allopolyploids are more frequently reported, the extent of autopolyploidy in plants is likely to be highly underestimated (Soltis et al. 2007); and where lineages present several different cytotypes, unless hybridization is documented, the evolutionary history of these lineages is likely to be obscured. Moreover, both autopolyploids and allopolyploids are likely to be of recurrent origin and a range of inheritance patterns may be observed (Soltis et al. 2007). Also, hybridization events may be multidirectional (Soltis 2000) and are often accompanied by further increases in ploidy, due to the production of unreduced gametes (Ramsey & Schemske 1998, Barket et al. 2016).

With respect to allopolyploids, in many cases, it can be difficult to recognize interspecific hybrids because of a high degree of similarity to one parent (Soltis et al. 2007) and this problem becomes increasingly difficult in groups with highly reduced morphology. Evidence here, along with previous studies (Les et al. 2010, Les et al. 2015), suggests that at least some of this chromosomal variation in *N. guadalupensis* potentially is of allopolyploid origin. Unfortunately, as Chase (1947) illustrated, chromosome size is similar in all three taxa precluding parental determinations from gross chromosome counts or flow cytometry; and alternative cytogenetic tools, such as FISH or GISH (Younis et al. 2015) would be required. Additionally, while traditional methods employ sequencing and cloning to unravel additive sequence variation in the nrITS region, these methods may not be sufficiently sensitive to detect any low copy repeats in older hybrids (Rauscher et al. 2002). Investigating the polyploid *Glycine tomentella* complex, where much of the diploid interspecific variation previously had been characterized, Rauscher et al. (2002) demonstrated how nrITS repeat-specific primers were an effective tool for detecting

evidence of recent and historical hybridization. Combining serial DNA dilutions of known diploid parents, they estimated that their primers were sufficiently sensitive to detect rare alleles at a ratio of approximately 1:1000 copies. By employing repeat-specific primers in this study, widespread introgression from the annual species (*N. canadensis* and *N. flexilis*) into *N. guadalupensis* has been detected.

Northern populations amplified both *N. canadensis* and *N. flexilis* alleles and this result was not entirely unexpected given the current sympatric ranges of these three species (Figure 1 and Figure 4 (B); along with previous evidence of hybridization between these taxa (Les et al. 2010). However amplification of *N. canadensis* alleles in the southern and widespread haplotypes, was not initially anticipated. Apart from the cG14 haplotype (in which only '*olivacea*' amplified the *N. flexilis* alleles), all other PCR amplifications from individuals in the widespread/southern haplotypes comprised *N. canadensis*, *N. guadalupensis* '*floridana*' or unknown divergent alleles that were most similar to the *N. canadensis* alleles.

Amplification of the '*floridana*' (nrG1) allele resulted from non-specific primer annealing, and primarily was limited to the known range of this subspecies in North America, as expected. Exceptions to this were a single accession from Wyoming (discussed below), and accessions representing a small extension of the known distributional range (Alabama and South Carolina).

However, *N. canadensis* alleles were found at great distances from the current distribution of this species (Figure 4 (B)). Nevertheless, given that the current ranges of both *N. canadensis* and *N. flexilis* are primarily above the last major glaciation boundary, these two taxa likely were further south during consecutive Pleistocene glaciations (Les et al. 2015); and a fossil

record for *N. flexilis* from Georgia, at 21 kybp (Watts, 1970), extends the historical range of this species considerably beyond its current distributional range.

While this study demonstrates hybridization and introgression from the two annual species into subsp. *guadalupensis*, to what degree either one or the other genomes has introgressed cannot be determined strictly from these markers, and the dosage effect of different ploidy levels also must be considered. With respect to the annual species, Chase (1947) found only diploids (*N. flexilis*) and tetraploids (*N. canadensis*) (Les et al. 2015) in his survey of northern populations of these two annuals. Yet, with the exception of a cross between *N. flexilis* (maternal) and *N. guadalupensis*, every hybrid combination involving these three species has been found in natural populations during the course of this North American *Najas* project. Chase (1947) clearly suspected hybridization and in particular drew attention to populations in Mirror and Kinderhook Lakes, New York where he noted "there are not only a remarkable assortment of species of *Najas* but also individuals which do not quite fit the species. For example, the seeds of *N. guadalupensis* from Mirror Lake plants tend to be much longer than expected and somewhat differently pitted (NY-54)… but judging from the known cases it is the higher polyploids of *N. guadalupensis* which are associated with *N. flexilis*".

While Chase found no triploids in his study to indicate hybridization between *N. canadensis* and *N. flexilis*, a number of hybrids, with multiple origins (and limited fertility) between the two annual species were identified during our study (Les et al. 2015). Clearly these three species have not diverged sufficiently to limit hybridization, to the extent that is now becoming evident; therefore many questions remain to be answered. How has genetic integrity been maintained in the two annual species and why has homogenization of the three taxa not occurred? Particularly when these three species are often found intermingled within a lake body.

What enhances or limits the success of any hybrids and what situations are likely to promote hybrid events?

Life History and sexual system in *Najas guadalupensis*

Different life history strategies will likely influence many different outcomes related to hybrid success, dispersal and establishment (Philbrick & Les 2009). Les et al. (2015) hypothesized that any gene flow from *N. guadalupensis* into *N. canadensis* or *N. flexilis* is likely to result in postzygotic barriers in these annual species (which require sexual reproduction for their continuity). On the other hand, clonal reproduction is long recognized as a mechanism to overcome such barriers associated with unbalanced polyploids (Grant 1971, Les & Philbrick 1993, Mallet 2007). If populations of *N. guadalupenis* are predominantly clonal, with limited sexual reproduction, this would allow individuals in these populations to harbor introgression from the two annual species.

Overlaid on all of this is the sexual system in *Najas*, which requires more investigation to determine whether or not it is plastic. While the two subgenera *Najas* and *Caulinia*, are separated as dioecious and monoecious species (Magnus 1870, Rendle 1899), Triest (1989) cautioned on making the assumption that all species within subgenus *Caulina* were strictly monoecious. Results here suggest that northern populations of *N. guadalupensis* are at least functionally dioecious. This raises the question whether there are certain ecological situations when female flowers are produced? Certainly, female flowers are necessary for any hybridization and introgression into *N. guadalupensis* to occur. Aquatic plants employ a diversity of strategies to ensure success in the aquatic environment and facilitate outcrossing when the opportunity arises. For example, a number of aquatic plants produce two different types of flowers, underwater

cleistogamous flowers (permanently closed flowers - self pollinated), and chasmogamous flowers (open flowers - promoting outcrossing), at the margins of water bodies. Two examples are *Glossostigma cleistanthum* W.R. Barker (Les et al. 2006) and *Ottelia alismoides* (L.) Pers. (Cook & Urmi-Konig 1984).

Could the sexual system in *N. guadalupensis* be plastic to environmental conditions? Is it possible that female flower production might be initiated in situations where plants are growing at marginal areas of water bodies and be triggered by changes in light, temperature or other factors? If, in these situations, resources were switched to sexual rather than vegetative production (Friedman & Rubin 2015) then, provided that complete drying out of marginal areas did not occur, the resulting seed progeny (or fragments with seed) could be dispersed back into the lake with water currents. A further consequence of this might be the promotion of outcrossing (if different species or genotypes were intermingled at lake edges). What might be the result on overall population and species dynamics of rare events like this? While undertaking field work at Wononskopomuck Lake, Connecticut in 2010, all three *Najas* species, along with hybrids between *N. canadensis* and *N. flexilis* were collected from a boat launch puddle at the edge of the lake. These small plants were intermingled and growing in less than 8 cm of water. How stressful conditions like this might induce flowering in *Najas* is not known, but certainly warrants further investigation. Obtaining the correct conditions for growing the annual cold water species *N. canadensis* and *N. flexilis* can be quite challenging; however, *N. guadalupensis* might be a more tractable plant for such experiments.

Within different life cycle patterns, differential strategies with respect to investment towards seed or vegetative reproduction will also influence dispersal, in terms of distance and establishment success. The low seed output of northern populations raises questions about how

these plants are being distributed. While seed propagules are likely to be dispersed further than asexual propagules (Figuerola & Green 2002), the wide scale distribution of many invasive aquatic plants through vegetative propagules attests to the success of this strategy in the aquatic environment. For example, one of the most serious aquatic weeds in the U.S. is *Hydrilla verticillata* L.f. Royle, first introduced to Florida in the 1950's (Schmitz et al. 1991) as a dioecious (female) clone (Benoit 2011). Likewise, the invasive Canadian pondweed (*Elodea canadensis* Michx.) was introduced to Britain as a single female clone (Arber 1920 p. 55).

Hybridization with the more cold tolerant annual species, *N. canadensis* and *N. flexilis*, may have supplied an important adaptive advantage to the southern species *N. guadalupensis*, by introgressing a level of cold tolerance into the *N. guadalupensis* genome. In 1971, Wentz & Stuckey documented the changing distribution of *Najas* in Ohio. One trend they identified was a reduction of the two *Najas* species associated with cooler, clear water (*N. flexilis* and *N. gracillima*), and simultaneous increase of three other *Najas* species, one of which was *N. guadalupensis*. Additionally, other authors have documented locally aggressive populations in New England (Hellquist 1977, Les et al. 2010). Again, various factors may contribute to the success of populations (including lake management regimes) but the ability of *N. guadalupensis* to overwinter vegetatively, presumably would give this species a competitive advantage over annual species like *N. canadensis* and *N. flexilis* seedlings in the early spring, and may be further responsible for the spread of this taxon at northern latitudes.

Clearly, many aspects of life history strategies in *N. guadalupensis* populations in North America, along with the interaction of this species with the annual *N. canadensis* and *N. flexilis*, require further investigation.

Najas guadalupensis in South America

With our focus on North American *Najas*, the extent of *N. guadalupensis* diversity in the Neotropics must not be overlooked. How North American *N. guadalupensis* populations interact with diversity in Central America and the Caribbean must also be considered. Given that major North American bird flyways funnel into these two regions, with several species of waterfowl overwintering in Central America and the Caribbean (Nichols et al. 1995, McGowan et al. 2007, Buhnerkempe et al. 2016), these regions are likely to be a zone of high *Najas* diversity. Perhaps it is not surprising therefore that populations in Texas (along with California, also a region for overwintering waterfowl) harbor the highest haplotype diversity in this study. Additionally, Florida represents the North American state with the greatest number of South American taxa within section *Americanae* (*N. filifolia*, *N. guadalupensis* '*floridana*', *N. guadalupensis* subsp. *guadalupensis* and *N. wrightiana*) (Lowden 1986).

In resolving misidentifications of *N. wrightiana* from The Bahamas, Lowden (1986) suggested that hybridization potentially occurred between *N. wrightiana* and *N. guadalupensis.* Also, in determining specimens from Charles Wright's Cuban collection [Wright 3716: NY1625868], Lowden annotated a specimen as "*N. wrightiana* with a tendancy towards guad". Although our limited sampling of *N. wrightiana* shows that it is genetically distinct from *N. guadalupensis*, *N. wrightiana* also shows a range of phenotypic variation and so it would be presumptive to draw any definitive conclusions here.

Recently, Ito et al. (2017) published a phylogeny of *Najas*. In the clade of interest to this study, they included three *N. guadalupensis* accessions, along with single accessions of *N. filifolia*, *N. wrightiana*, *N. canadensis*, *N. guadalupensis* '*muenscheri*' [= *N. canadensis*] and *N.*

flexilis. All of these accessions represented sequences previously published (Les et al. 2010, 2015), with the exception of two novel sequences of *N. guadalupensis*, one from cultured material introduced to Japan (Ito Y. 1142 & al.; TNS), and the other from Cordoba, Argentina (Ito Y. 1996 & al.; TNS).

Ito et al. (2017) also experienced conflict between their chloroplast (*matK, rbcL, rpoB, rpoC1*) and nrITS phylogenies, with respect to the position of the three *N. guadalupensis* samples. While making no reference to polymorphic sequences for *N. guadalupensis*, Ito et al. also experienced difficulty sequencing the nrITS region in their two *N. guadalupensis* samples, with 79.5 % of this region coded as missing data or gaps in the Japanese sample and 14% in the Argentinean sample. Unfortunately, the *matK* region sequenced in their study did not correspond to or overlap with the *trnK*/*matK* region sequenced here and the *rbcL* region is not sufficient for resolution. However, a quick Bayesian analysis (GTR model) of their nrITS sequences, combined with the *N. guadalupensis* nrITS ribotypes from this study (data not shown), placed the Argentinean sample in the clade with '*floridana*' (nrG1), and the Japanese sample as sister to the rest of the *N. guadalupensis* group. With such a large amount of missing data, it would have been of interest had the authors subcloned their *N. guadalupensis* sequences, especially as chromosome counts were provided for the Japanese specimen (*2n* = ca. 48), which indicated that this specimen is of polyploid origin. Unfortunately, I was unable to access herbarium images for these two taxa.

In the current study, in addition to demonstrating admixture from *N. canadensis* and *N. flexilis*, sequences of unknown origin (nrUn), most closely related to *N. canadensis*, were also amplified from *N. guadalupensis*. With the exception of a single individual from southern Indiana, the unknown nrITS sequences in this analysis were only associated with

southern/widespread haplotypes. The accession from Indiana possessed the northern haplotype (cG16), but also possessed the widespread nrITS ribotype (nrG10), in combination with the northern ribotype (nrG12), indicating that this individual is an interracial hybrid in this zone of secondary contact between the widespread and northern/eastern haplotypes.

Without further elucidation of within-species diversity in Central America, the Caribbean, and South America in general, it is difficult to interpret how these unknown alleles fit into overall *N. guadalupensis* genetic diversity. All of these sequences had conserved 5.8s motifs (Harpke & Peterson 2008), unlike the divergent *N. flexilis* sequences which are interpreted here as pseudogenes. Therefore it is impossible to ascertain whether these sequences represent pseudogenized *N. canadensis* alleles or other, previously unsampled, genetic diversity within *N. guadalupensis* (or another unsampled member of section *Americanae*).

In reflection of the words of Magnus (1870), further molecular sampling may reveal that *N. guadalupensis* might be considered a compilospecies *sensu* Harlan & de Wet (1963), actively taking up genetic material from congeners when they co-occur.

***Najas guadalupensis* "*floridana*"**

Without sufficient sampling of populations outside of North America it is difficult to draw any firm conclusions with regards to '*floridana*'. In North America, this taxon is only known from the states of Florida and Georgia; however, it is distributed throughout the Caribbean region with a single account from Guatemala in Central America (Lowden 1986, Figure 1). Although not recovered in our sampling subsp. *guadalupensis* is sympatric with '*floridana*' in the Florida region (Lowden 1986, Haynes 2000). All of our accessions from Florida and Georgia with the '*floridana*' phenotype possessed a distinct nrG1 ribotype, which resolved outside of the subsp.

guadalupensis, *N. canadensis* and *N. flexilis* clades; however, the chloroplast haplotype (cG7) of these accessions associated with the widespread subsp. *guadalupensis* clade (cG9–cG12). Additionally, haplotype diversity in the trnLF region within '*floridana*' was greater in this restricted geographic sample, compared with that in the other subsp. *guadalupensis* clades.

Reproductive output in our '*floridana*' accessions was very low, which supports a hybrid origin. Only four seeds in total were found on three of our 12 accessions; and while male flowers were present on one sample, seven accessions were completely sterile. Davenport (1980) obtained a chromosome count of *2n* = 48 from a single '*floridana*' individual from Florida, which also lends support to the supposition that our sampled populations of '*floridana*' are of hybrid origin.

Our lack of sampling throughout the distribution of this taxon (Figure 1), precludes us from determining whether North American populations of this taxon represent clonal hybrids between *N. guadalupensis* and a now extirpated species; or whether a taxon with a similar '*floridana*'phenotype, harboring the nrG1 ribotype, along with a "pure" '*floridana*' haplotype still exists inside or outside of North America. It should be noted that all extant South American species of section *Americanae* have exerted teeth margins, as does '*floridana*'.

Additionally, the account of a single individual with the '*floridana*' phenotype in Guatemala is puzzling (Lowden 1986). Lowden was extremely careful in documenting his determinations and observations, in many cases drawing attention to incorrect determinations and mixed herbarium sheets. We can therefore assume that this report is an accurate distributional record. Whether more populations corresponding to this phenotype exist in Central America is unknown.

Nomenclature

As this study deals with *N. guadalupensis* in North America, for clarity, I have followed the subspecific epithet of Haynes (2000) in the Flora of North America; however, I believe nomenclatural priority was not observed in naming '*floridana'*. Braun (1864) had already applied the name *N. flexilis* var. *curassavica* to a collection of this taxon from El Hato, Curaçao, in the Lesser Antilles. Later, in 1868, he revised this name to *N. microdon* [=*N. guadalupensis*] var. *curassavica* A. Br. In 1870, Magnus (Braun's student), provided a drawing of this specimen, which clearly shows the marginal leaf spines, subtended by a number of cells (Plate V: Figure 19). Later authors (e.g., Rendle 1899, Chase 1947) and Clausen (1936), also recognized this morphology from populations in Florida and applied Braun's name. In his revision of *Najas*, Haynes (1979) listed *N. microdon* var. *curassavica* as a synonym of *N. guadalupensis* var. *guadalupensis*, yet a review of available online herbarium specimens reveals a specimen determined by Clausen in 1938 as *N. guadalupensis* var. *curassavica* (MICH1432792), which was later redetermined by Haynes as *N. guadalupensis* var. *floridana*.

Until more material can be sampled from Central America and the Caribbean, it is difficult to make any definitive statement with regards to '*floridana'*. Given the distinct phenotype, I believe that it should be retained at the rank of subspecies. Should future research recover a genetically distinct maternal haplotype, and fertile populations, it may be preferable to elevate this taxon to species level. However, whether it is retained at subspecies level or elevated, I believe that any decision should take into account Braun's nomenclatural priority.

***Najas guadalulpensis* "olivacea"**

The original description of *N. olivacea* (Rosendahl and Butters 1935) was from a population in Norway Lake, Minnesota (now extirpated). Currently however, in the Flora of North America, subsp. *olivacea* is recorded as having an extensive northern distribution (Haynes 2000, Figure 1), associated with the Great Lakes Region and extending as far north as Manitoba, Ontario and Quebec (with an isolated record for Massachusetts).

In our analysis, the type material of *N. olivacea* is represented by the cG14 *rbcL*/*trnK*/*matK* haplotype. Individuals in this clade have the same, more slowly evolving *rbcL* sequence as populations in the widespread haplotype clade, but unlike individuals in that clade, which predominantly associated with the widespread ribotypes (nrG7–nrG11), all individuals within the cG14 haplotype possessed the northern ribotypes (nrG12 or nrG14). It seems likely, therefore, that this clade represents populations in a zone, associated with the Great Lakes, where *N. guadalupensis* populations radiating postglacially from the Mississippi embayment region are coming in to secondary contact with Atlantic coastal state populations.

Two distinct groups were evident within this haplotype. One group associated with the '*olivacea*' phenotype, and possessed the nrG14 ribotype, along with evidence of introgression from *N. flexilis*. All accessions represented by this combination are restricted to Minnesota and northwestern Wisconsin. Apart from the type material for '*olivacea*', seed was only present on one other accession. This accession also possessed seed with intermediate characters between *N. guadalupensis* and *N. flexilis* (Figure 4 (D), Appendix G). Combined with triploid (*2n*=36) chromosome counts (Chase 1947), all evidence here supports a hybrid hypothesis, a supposition originally proffered by Chase.

Our other accessions that shared this cG14 haplotype were restricted to northwestern Pennsylvania. These individuals were highly similar to each other vegetatively, and superficially similar to '*olivacea*' in having short, abruptly acute leaves. However, all of these plants were completely sterile, did not have the thickened stems of '*olivacea*' (appendix H) and possessed the nrG12 ribotype, with admixture from *N. canadensis* rather than *N. flexilis*.

The final step in the key for *N. guadalupensis* in the Flora of North America (Haynes 2000) separates '*olivacea*' from subsp. *guadalupensis* by the number of leaf margin teeth and stem diameter. More accurate evaluation of leaf tooth number for *N. guadalupensis*, in general, needs to be established, as will be discussed below. That aside, unfortunately Haynes (1979) failed to record any voucher specimens in his revision of the genus in North and Central America, making it difficult to track his determinations and understand how he evaluated the distribution of this taxon. A review of online digital images leads me to conclude that several of Haynes' determinations, with respect to '*olivacea*', should be re-evaluated. Results here suggest that '*olivacea*', as originally described, has a much more restricted distribution west of the Great Lakes. Specimens from Ontario (MINN1189338, MINN1189343), which Haynes has determined as subsp. *olivacea*, clearly do not correspond to his own description of this taxon. Additionally, in 1936, Clausen described a collection of 'peculiar *Najas*' by Muenscher and Lefler (18239) from Cayuga Lake, NY, but was hesitant to definitively call it *N. olivacea*, as the material was "absolutely sterile". Clausen described these specimens as slender with a tooth count of 50-75 teeth, yet Haynes later determined Clausen and Muenscher's specimens from Cayuga Lake as subsp. *olivacea*, which using his own key, should result in subsp. *guadalupensis* instead (IND0009431, MINN1189344, MINN1189336, TENN-V-0013888).

In conclusion, the population on which the original description for *N. olivacea* was based is now extirpated and results here suggest that this hybrid form is much more restricted and does not have the extensive range attributed to it in the Flora of North America (Haynes 2000). Given our results, this taxon likely represents only one of a series of hybrid populations between *N. guadalupensis* and *N. flexilis* or *N. canadensis*. A variety of localized, predominantly vegetative forms likely exist in northern *N. guadalupensis*, and it is suggested that the subspecific status of '*olivacea*' no longer is warranted.

Additional notes

Rice culture – Butte County, California

The occurrence of five accessions in the cG13 clade representing a highly disjunct distribution between the state of Mississippi and Butte County, California is curious, and a possible explanation for this discontinuity may involve rice culture in North America. In the mid to late 1800s the principal portion of the U.S. rice crop came from the southern Atlantic and Gulf states (Knapp 1900) with commercial rice production in California beginning later, in 1912 (Wilson 1979). Currently the states of Arkansas, California, Louisiana, Mississippi and Texas represent the five largest rice-producing states in the United States (National Agricultural Statistics Service 2016). Although regarded as native to California (Thorne et al. 2017), *N. guadalupensis* presently is associated with rice cultivation there, where it has become a troublesome weed in man-made ponds and disturbed or controlled aquatic systems (DiTomaso & Healy, 2003). Early accounts of *N. guadalupensis* in California (Los Angeles and San Bernadino) date back to the 1890s, and yet the first account of *N. guadalupensis* from Butte County is as recent as 1946 (data provided by the participants of the Consortium of California Herbaria: http://

ucjeps.berkeley.edu/consortium/). This record is from an irrigation ditch at Greylodge Game Refuge, approximately 10 miles from Richvale, the birthplace of California rice (Lee 2005). Notably, on the following day, the same collector, Herbert Mason, provided the first Californian record of *N. graminea* Delile, an introduced Asian species and known rice seed contaminant (Triest 1988). Similarly, *Monochoria vaginalis* (Burm. f.) C. Presl ex Kunth. and *Heteranthera limnosa* (Sw.) Willd. were introduced to this area around this time (Barrett 1993), as rice seed contaminants. Butte County, California is represented by three haplotypes in our analysis, with the state of California recording some of the highest haplotype diversity. Given that many *Najas* species are regarded as rice field weeds (Triest 1988), the possibility that some of these Californian populations may represent non-indigenous genotypes associated with more recent introductions cannot be ruled out.

Fish rearing and aquarium trade - Indiana, Louisiana, Oklahoma, Wyoming

As mentioned, *N. guadalupensis*, or guppy grass, has had a long association with the fish rearing trade, resulting in invasive populations being reported far outside of their natural range, for example, in Hawaii (Staples et al. 2000) and Japan (Ito et al. 2017). Some anomalous genotypes in our study, point to the possibility of human-mediated introductions of *N. guadalupensis* in North America potentially associated with fish rearing for lake stocks, or the tropical aquarium trade.

The presence of a northern (nrG16) haplotype in an accession from southern Indiana, along with two accessions in Louisiana and Oklahoma possessing northern ribotypes (nr12), indicate possible non-indigenous introductions through the fish rearing trade. One sample was collected at Dream Lake, Indiana, a recreational fishing lake, that was only constructed in 1965

(http://www.in.gov/dnr/forestry/4825.htm); another sample was collected by a colleague at Louisiana State Aquaculture Research Station, Denton County, Louisiana, an experimental station for freshwater fish and crawfish; and the third accession came from Lake Murray, Oklahoma, which is part of the largest public park in that state and one of the top 100 amenity and fishing lakes in the U.S. (http://www.lake-murray.org).

Similarly, a presence of the '*floridana*' ribotype (nrFL1) in an accession from Wyoming raises questions. Not only is this collection well outside of the range of '*floridana*', but it also represents a new state record for *N. guadalupensis*. This sample was collected from Kelly Warm Springs, Grand Teton National Park by a colleague; and it was noted at the time of collection that a number of introduced fish and mollusks were present (B. Hellquist, pers. comm.). Apparently, this warm spring has been an illegal dumping ground for aquarium tropical fish, with records, dating back to the 1940s, of guppies being introduced (https://parkplanning.nps.gov/projectHome.cfm?projectID=59971). It is likely therefore that *N. guadalupensis* has been introduced (with the contents of an aquarium) to this waterbody.

Although the distribution of *N. guadalupensis* through the registered trade has waned in recent years, a recent search on Ebay® for guppy grass (https://www.ebay.com) reveals that *N. guadalupensis* is still passing through private hands in the U.S. This study further cautions that more regulation of the unregistered sale of aquatic plants through the internet is necessary.

Turion production

Perennial water plants excel in producing a variety of structures for vegetative propagation and survival under adverse conditions (Philbrick & Les 1996). Along with the many vegetative structures associated with propagation in terrestrial plants (e.g., corms, stolons, rhizomes and

gemmiparous plantlets), in addition, aquatic plants may fragment from lateral growth containing individual nodes (or seed) capable of regeneration in the water environment. Like terrestrial plants, dormancy may be achieved in different ways e.g., seed, corms, rhizomes or by winter buds or highly specialized buds called turions. However, unlike in terrestrial plants, in the aquatic environment these buds may detach from the parent plant, remain viable and become highly vagile allowing dispersal over great distances (Philbrick & Les 1996). Indeed, the ability of these buds to detach from parent plants led Hutchinson (1975) to coin the term "asexual annuals". Such perennating structures also inhibit effective eradication of invasive clonal aquatic plant plants, for example *Hydrilla*.

In 1939 Rosendahl reported winter buds in '*olivacea*' and greenhouse experiments showed that these buds propagated readily when broken off. Winter buds were also reported from *N. horrida* (Triest 1988), and turions were reported for populations of *N. marina* in Israel (Agami & Waisel 1986). Recent evidence suggests that northern populations of *N. guadalupensis* may also be capable of producing turions. Plants collected from Mansfield Hollow, Connecticut, in September, 2017 revealed the presence of a highly modified bud enclosed in the leaf sheets. This ovate bud had closely packed leaves, and appeared more like a *Najas* seed than a normal shoot tip (Appendix I).

As a number of collections of *N. guadalupensis* for this study were indicated as aggressive or invasive [NH, WA, WI], it is advised that lake managers and collectors of *N. guadalupensis* should be vigilant for these structures, to determine whether these buds are effective overwintering and dispersal structures.

Leaf marginal tooth number

The first report of the upper limit of 100 teeth per leaf margin side for *N. guadalulpensis* appears to have originated from Morong (1885) who reported a count of 30-100 teeth (Table 3). Later, without stating a reason, he revised this count to between 40-50 teeth (Morong 1893). In their original description of var. *floridana*, Haynes and Wentz (1974) used leaf tooth number as one of the characters distinguishing var. *guadalupensis* and var. *floridana*, and at that time they reported a count of ca. 100 teeth for var. *guadalupensis*. In the Flora of North America, Haynes (2000) then revised this count to 50-100 teeth for subsp. *guadalupensis*. The original type specimen for *N. guadalupensis* from Guadaloupe was destroyed during World War II and Haynes and Wentz (1974), expressing uncertainty about which variety Sprengel had before him when he originally described *Caulinia guadalupsis* [=*N. guadalupensis*], stated that having examined material from the type locality "we consider that taxon as the typical variety". Furthermore, their decisions were based on their "own field work and the examination of several thousand specimens" of *N. guadalupensis*. Interestingly, earlier Braun (1864) had recorded a leaf tooth count of "about 20 teeth" from Guadaloupe (Duchaissing, MO), which specimen Lowden (1986) also determined as *N. guadalupenis* forma *guadalupensis*. Unfortunately Lowden (1986) provided no details of leaf tooth number in his revision for *N. guadalupensis*.

While a comprehensive study of leaf teeth number was not undertaken during this study, marginal teeth were counted on a number of specimens and all were within the range found by previous researchers, with none reaching the upper limits reported by Haynes (Table 3).

Table 3. Summary of leaf tooth counts in *N. guadalupensis* subsp. *guadalupensis* from various authors.

Author	Tooth number	Localities
Braun 1864	ca. 20	Guadaloupe
Schumann 1894	27-35	Brazil
Rendle 1899	20-45	Southern US, Central and South America and Caribbean
Morong 1885	30-100	Texas
Morong 1893	40-50	Southern US, Caribbean
Wentz & Haynes 1973	ca. 100	Panama
Haynes & Wentz 1974 (var. *guadalupensis*)	ca. 100	North America, Panama, Guadaloupe
Haynes & Holm-Nielsen 1986	30-100	Ecuador
Haynes 1979 (var. *guadalupensis*)	50-100	North America

Collection advice

Lowden (1986) first alerted future workers to the possibility of having more than one *Najas* taxon on a herbarium sheet and he spent considerable effort in documenting mixed herbarium sheets, along with mixed specimens under the one collection number that had been distributed to various herbaria.

In the earlier stages of our study we were also alerted to the possibility of having a number of taxa included in a single collection. The stems of *Najas* are highly brittle and even with the greatest of care, ensuring extraction of a single plant through the water interface can be challenging. With such highly reduced morphology and overall similarity of several species of *Najas*, I think it is worth restating the cautionary advice of Lowden again here.

Preferably specimens should be floated out prior to drying, to disentangle them, and secured to herbarium sheets with tape rather than herbarium glue. This offers the best opportunity to preserve *Najas* as a scientifically useful herbarium specimen.

References

Agami, M. Beer, S. & Y. Waisel. 1986. The morphology and physiology of turions in *Najas marina* L. in Israel. Aquatic Botany 26: 371–376.

Álvarez, I. & J.F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29: 417–34.

Amezaga, J.M. Santamaria, L. & Green, A.J. 2002. Biotic wetland connectivity – supporting a new approach for wetland policy. Acta Oecologica – International Journal of Ecology 23, 213–222.

Anderson, L.W.J. 1990. Aquatic weed problems and management in the western United States and Canada. In: Pieterse AH and Murphy KJ (eds) AquaticWeeds pp 371–391. Oxford University Press, Oxford

Arber, A. 1920. Water Plants: A Study of Aquatic Angiosperms. Cambridge University Press. Cambridge.

Barbier, P. 1989. Genetic variation and ecotypic differentiation in the wild rice specis *Oryza rufipogon*. II. Influence of the mating system and life-history traits on the genetic structure of populations, Jpn. J. Genet. 64: 273–285.

Barker, M.S. Arrigo, N. Baniaga, A.E. Li, Z. & D.A. Levin. 2016. On the relative abundance of autopolyploids and allopolyploids. New Phytologist 210: 391–398.

Benoit, L.K. 2011. Cryptic speciation, genetic diversity and herbicide resistance in the invasive aquatic plant *Hydrilla verticillata* (L.f.) Royle (Hydrocharitaceae). Ph.D. dissertation, University of Connecticut, Storrs, CT. Website (http://digitalcommons.uconn.edu/dissertations/AAI3492149).

Birks, H.H. 2006. Plant Macrofossils. In: Last W.M. and Smol J. P. (eds) Tracking Environmental Change Using Lake Sediments: Volume 3: Terrestrial, Algal, and Siliceous Indicators. Edition: Developments in Paleoenvironmental Research Vol.3, pp.49-74. Kluwer, Dordrecht, The Netherlands.

Braun, A. 1864. Revision of the genus *Najas* of Linnaeus. Seemann's Journal of Botany 2: 274–279.

Bruen, T.C. Philippe, H. & D. Bryant. 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172: 2665–2681.

Buehler, D.M, Baker, A.J & T. Piersma. 2006. Reconstructing palaeoflyways of the late Pleistocene and early Holocene red knot (*Calidris canutus*). Ardea 94: 485–498.

Buhnerkempe, M.G. Webb, C.T. Merton, A.A. Buhnerkempe, J.E. Givens, G.H. Miller, R.S. & J.A. Hoeting. 2016. Identification of migratory bird flyways in North America using community detection on biological networks. Ecological Applications 26:740–751.

Cain, M.L. Milligan, B.G. & A.E. Strand. 2000.Long-distance dispersal in plant populations. American Journal of Botany. 87: 1217–1227.

Campbell, D. 1897. A morphological study of *Naias* and *Zannicheliia*. Proc. Calif. Acad.Sc. 1:1–62.

Campbell, C.S. Wojciechowski, M. F. Baldwin, B.G. Alice, L.A. & M.J. Donoghue. 1997. Persistent nuclear ribosomal DNA sequence polymorphism in the Amelanchier agamic complex (Rosaceae). Mol. Biol. Evol. 14: 81–90.

Chase, S.S. 1947. Preliminary Studies in the Genus *Najas* in the United States. Ph.D. dissertation, Cornell University, Ithaca, New York.

Clausen, R.T. 1936. Studies in the genus *Najas* in the northern United States. Rhodora 39: 334–345.

Clausen, R. T. 1937. A new species of *Najas* from the Hudson River. Rhodora 39: 57–60.

Cook, C.D.K. & K. Urmi-Konig. 1984. A revision of the genus *Ottelia* (Hydrocharitaceae). 2. The species of Eurasia, Australasia and America. Aquat. Bot. 20: 131–177.

Davenport, L.J. 1980. Chromosome number reports LXVII. Taxon 29: 351.

de Candolle A. 1855. Geographie botanique raisonnee; ou, exposition des faits principaux et des lois concernant la distribution geographique des plantes de l'epoque actuelle. 2 vols. Victor Masson, Paris. 1365 pp.

DiTomaso J.M. & Healy E.A. 2003. Aquatic and riparian Weeds of the West. University of California Agriculture and Natural Resources, Publication 3421, Berkeley, California, pp. 107–108.

Dieffenbacher-Krall, A.C., & G.L. Jacobson J.R. 2001. Post-glacial changes in the geographic ranges of certain aquatic vascular plants. Proceedings of the Royal Irish Academy 101: 79–84.

Doyle, J.J. & Doyle, J.L. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19: 11–15.

Figuerola, J. & A.J. Green. 2002. Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies. Freshwater Biol. 47: 483–494.

Fox, G.A. 1990. Perennation and the Persistence of Annual Life Histories. The American Naturalist 135: 829–840.

Friedman, J. & M.J. Rubin. 2015. All in good time: understanding annual and perennial strategies in plants. American Journal of Botany 102: 497–499.

Gaut, B. Yang, L. Takuno, S. & L.E. Eguiarte 2011.The patterns and causes of variation in plant nucleotide substitution rates. Annual Review of Ecology Evolution and Systematics 42: 245–266.

Grant V. 1971. Plant speciation, Columbia University Press, New York.

Harlan J.R. & J.M.J. de Wet. 1963. The compilospecies concept. Evolution 17: 497–501.

Harpke D. & A. Peterson. 2008. 5.8 S motifs for the identification of pseudogenic ITS regions. Botany 86: 300–305.

Haynes, R.R. 1979. Revision of North and Central American *Najas* (Najadaceae). Sida 8, 34–56.

Haynes, R.R. 2000. 197. Najadaceae Jussieu – Naiad or water-nymph family. In: Flora of North America Editorial Committee (eds.), Flora of North America North of Mexico, vol. 22, Magnoliophyta: Alismatidae, Arecidae, Commelinidae (in Part), and Zingiberidae. Oxford University Press, New York City, New York, pp. 77–83.

Hellquist C.B. 1997. A guide to invasive non-native aquatic plants in Massachusetts. Massachusetts Department of Environmental Management, Boston, Massachusetts. 15 pp.

Hutchinson, G.E. 1975. A Treatise on Limnology Volume 3. Limnological Botany. John Wiley & Sons, New York. 238 pp.

Ito, Y. Tanaka, N. Gale, S. Yano, O. & J. Li. 2017. Phylogeny of Najas (Hydrocharitaceae) revisited: Implications for systematics and evolution. Taxon, 66: 309–323.

Karrow, P.F. Bloom, A.L. Haas, J.N. Heiss, A.G. McAndrews, J.H. Miller, B.B. Morgan, A.V. & K.L. Seymour. 2009. The Fernbank interglacial site near Ithaca, New York, USA. Quatern. Res. 72: 132–142.

Kasselmann, C. 2003. Aquarium Plants. Krieger Publishing Company, Malabar, Florida. 371 pp.

Knapp, S.A.1900. Rice culture in the United States (Public domain ed.). U.S. Department of Agriculture. pp. 6.

Lee C. 2005 The Birthplace of California Rice. Califonia Country Magazine. Accessed online: http://californiabountiful.com/features/article.aspx?arID=234.

Les, D. H. 1986. The phytogeography of *Ceratophyllum demersum* and *C. echinatum* (Ceratophyllaceae) in glaciated North America. Canadian Journal of Botany 64: 498–509.

Les D.H. 1988. Breeding systems, population structure, and evolution in hydrophilous angiosperms. Annals of the Missouri Botanical Garden 75: 819–835.

Les, D.H. & C.T. Philbrick. 1993. Studies of hybridization and chromosome number variation in aquatic plants: evolutionary implications. Aquatic Botany 44: 181–228.

Les, D.H. & L.J. Mehrhoff. 1999. Introduction of nonindigenous aquatic vascular plants in southern New England: a historical perspective. Biological Invasions 1: 281–300.

Les, D.H. Capers, R.S. & N.P. Tippery. 2006. Introduction of *Glossostigma* (Phrymaceae) to North America: a taxonomic and ecological overview. American Journal of Botany 93: 927–939.

Les, D.H. Sheldon, S.P. & N.P. Tippery. 2010. Hybridization in hydrophiles: natural interspecific hybrids in *Najas* (Hydrocharitaceae). Syst. Bot. 35: 736–744.

Les, D.H. Peredo, E.L. Benoit, L.K. Tippery, N.P., King, U.M. & S.P. Sheldon. 2013. Phytogeography of *Najas gracillima* (Hydrocharitaceae) in North America and its cryptic introduction to California. American Journal of Botany 100: 1905–1915.

Les, D.H. Peredo, E.L. Tippery, N.P. Benoit, L.K. Razifard, H. King, U.M. Na, H.R. et al. 2015. *Najas minor* (Hydrocharitaceae) in North America: A reappraisal. Aquatic Botany 126: 60–72.

Les, D.H. Peredo, E.L. King, U.M. Tippery, N.P. Ball, C. & R. Shannon. 2015. "Through thick and thin" : cryptic sympatric speciation in the submersed genus *Najas* (Hydrocharitaceae). Molecular Phylogenetics and Evolution 82: 1–53.

Lowden, R. M. 1986. Taxonomy of the genus *Najas* L. (Najadaceae) in the Neotropics. Aquatic Botany 24: 147–184.

Magnus, P. 1870. Beiträge zur Kenntniss der Gattung *Najas*. Berlin: L. G. Reimer.

Mallet J. 2007. Hybrid speciation. Nature 446: 279–283.

Martin A.C & F.M. Uhler. 1939. Food of game ducks in the United States and Canada. U.S. Department of Agriculture Technical Bulletin No. 634. Government Printing Office, Washington, D.C., USA.

Malinska H. Tate J.A, Matyasek R, Leitch A.R, Soltis D.E. et al. 2010. Similar patterns of rDNA evolution in synthetic and recently formed natural populations of *Tragopogon* (Asteraceae) allotetraploids. BMC Evol Biol. 10: 291.

McGowan, A. Woodfield, N. K. Hilton, G. Broderick, A. C. & B.J. Godley. 2007. A rigorous assessment of the avifauna of a small Caribbean Island: A case study in Anegada, British Virgin Islands. Caribbean Journal of Science. 43: 99–116.

Moody M.L. & Les D.H. 2007. Geographic distribution and genotypic composition of invasive hybrid watermilfoil (*Myriophyllum spicatum* x *M. sibiricum*) populations in North America. Biological Invasions 9: 559–570.

Morong, T. 1893. The Naiadaceae of North America. Mem. Torrey Bot. Club 3: 1–65.

Műhlberg, H. 1982. The Complete Guide to Water Plants. EP Publishing Ltd.Leipzig. 308 pp.

National Agricultural Statistics Service 2016. United States Department of Agriculture. All rice 2016 harvested acres by county for selected states. Website: https://www.nass.usda.gov/Charts_and_Maps/graphics/AR-PR-RGBChor.pdf. Accessed on: 15th September 2017.

Neiva, J. Serrão, E. A. Anderson, L. Raimondi, P. T. Martins, N. Gouveia, L. et al. 2017. Cryptic diversity, geographical endemism and allopolyploidy in NE Pacific seaweeds. BMC Evolutionary Biology 17: 30.

Nichols, J.D. Johnson, F.A. & B. K. Williams. 1995. Managing North American waterfowl in the face of uncertainty. Annual Review of Ecology and Systematics. 26:177–199.

Philbrick, C.T. & D.H. Les. 1996. Evolution of aquatic angiosperm reproductive systems. Bioscience 46: 813–826.

Ramsey J & D.W. Schemske. 1998. Pathways, mechanisms, and rates of polyploidy formation in flowering plants. Annual Review of Ecology and Systematics 29: 467–501.

Rauscher, J.T. Doyle, J.J. & H.D. Brown. 2002. Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex. Mol. Ecol. 11: 2691–2702.

Rendle, A.B. 1899. A systematic revision of the genus *Najas*. Trans. Linn. Soc. London, Bot. 5: 379–436.

Rogstad, S.H. 1992. Saturated NaCl-CTAB solution as a means of field preservation of leaves for DNA analyses. Taxon 41: 701–708.

Rosendahl, C.O. & F.K. Butters. 1935. The genus *Najas* in Minnesota. Rhodora 37: 345–348.

Rosendahl, C. 1939. Additional notes on *Najas* in Minnesota. Rhodora 41:187–189.

RStudio Team 2016. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Rüegg, S. U. Raeder, A. Melzer, G. Heubl, & C. Bräuchler. 2016. Hybridisation and cryptic invasion in *Najas marina* L. (Hydrocharitaceae)? Hydrobiologia 784: 381–395.

Santamaría, L. 2002. Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment. Acta Oecologia 23: 137–154.

Sawada, M. Viau, A.E. & K. Gajewski 2003. The biogeography of aquatic macrophytes in North America since the Last Glacial Maximum. Journal of Biogeography 30: 999–1017.

Schuman, K M. 1894.Najadaceae in K. F. Martius (eds.) Flora Brasiliensis Vol III Pars III pp716-735.

Sculthorpe, C.D. 1967. The biology of aquatic vascular plants. London: Edward Arnold.

Shafer, A.B.A. Cullingham, C.I. Cote, S.D. & D.W. Coltman. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Molecular Ecology 19: 4589–4621.

Schmitz, D.C. Nelson, B.V. Nall, L.E. & J. D. Schardt. 1991. Exotic aquatic plants in Florida: a historical perspective and review of the present aquatic plant regulation program, pp. 303-326. *In*: T.D. Center, R.F. Doren, R.L. Hofstetter, R.L. Myers, and L.D. Whittaker, eds., Proceedings

of the Symposium on Exotic Pest Plants, United States Department of the Interior, National Park Service, Washington, D.C.

Soltis, P.S. & D.E. Soltis. 2000. The role of genetic and genomic attributes in the success of polyploids. Proceedings of the National Academy of Sciences, USA 97: 7051–7057.

Soltis, D. E. Soltis, P. S. Schemske, D. W. Hancock, J. F. Thompson, J. N. Husband B.C. & W.S. Judd 2007. Autopolyploidy in angiosperms: Have we grossly underestimated the number of species? Taxon 56: 13–30.

Soons, M.B. van der Vlugt, C. van Lith, B. Heil, G.W. & M. Klaassen 2008. Small seed size increases the potential for dispersal of wetland plants by ducks. Journal of Ecology 96: 619–627.

Soria-Hernanz, D.F. Fiz-Palacios, O. Braverman, J.M. Hamilton, M.B. 2008. Reconsidering the generation time hypothesis based on nuclear ribosomal ITS sequence comparisons in annual and perennial angiosperms. BMC Evol. Biol. 8: 344.

Staples, G.W. Herbst, D.R. & C.T. Imada 2000. Survey of invasive or potentially invasive cultivated plants in Hawai‘i. Bishop Museum Occasional Papers. 651–31.

Sukumaran, J. & M.T. Holder. 2010. DendroPy: A Python library for phylogenetic computing. *Bioinformatics* 26: 1569–1571.

Thorne, R.F. Hellquist, C.B. & R.R. Haynes. 2017. *Najas guadalupensis subsp. guadalupensis*, in Jepson Flora Project (eds.) Jepson eFlora, http://ucjeps.berkeley.edu/cgi-bin/get_IJM.pl?tid=80549, accessed on September 1, 2017.

Triest, L. 1988. A revision of the genus *Najas* L. (Najadaceae) in the Old World. Mém. Acad. Roy. Sci. Outre-Mer, Cl. Sci. Nat. Méd. 22: 1–172.

Treist, L. 1989. Flora of Tropical East Africa - Najadaceae, Volume 125. A. A. Balkema, Rotterdam. p. 2

Untergasser, A. Cutcutache, I. Koressaar, T. Ye, J. Faircloth, B.C. Remm, M. & S.G. Rozen. 2012 Primer3 – new capabilities and interfaces. Nucleic Acids Research 40: 15.

van Kleunen, M. 2007. Adaptive genetic differentiation in life-history traits between populations of *Mimulus guttatus* with annual and perennial life-cycles. Evol Ecol. 21: 185–199.

Viau, A.E. Gajewski, K. Sawada M.C. & P. Fines. 2006. Millennial-scale temperature variations in North America during the Holocene. Journal of Geophysical Research, 111: p. D13102

Veit, R.R. 2000. Vagrants as the expanding fringe of a growing population. Auk 117:242–246.

Vesper, S.J. & R.L. Stuckey. 1977. The return of aquatic vascular plants into the Great Lakes region after Late-Wisconsin glaciations. In R. C. Romans [ed.], Geobotany: Proceedings of Geobotany Conference, 283–299. Plenum Press, New York, New York, USA.

Viinikka, Y. 2009. The occurrence of B chromosomes and their effect on meiosis in *Najas marina*. *Hereditas* 75: 207–212.

Watts, W.A. & R. Ross. 1960. Interglacial deposits at Kilbeg and Newtown, Co. Waterford. Proceedings of the Royal Irish Academy. 60:79–134.

Watts, W.A. 1970. The full-glacial vegetation of northwestern Georgia. Ecology 51: 17–33.

Wentz, W.A., & R.L. Stuckey. 1971. The changing distribution of the genus *Najas* (Najadaceae) in Ohio. Ohio Journal of Science 71: 292–302.

White, T. J. Brims, T. Lee, S., & J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In "PCR Protocols: A Guide to Methods and Applications" (M. Innis, D. Gelfand, J. Sninsky, and T. White, eds.), Academic Press, San Diego. pp. 315–322.

Wickham, H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Willson, J.H. 1979. Rice in California. In: Willson, J.H. (ed.). Rice in California. Butte County Rice Growers Association. pp. 20–40.

Younis, A. Ramzan, F. Hwang, Y.J. & K.B. Lim. 2015. FISH and GISH: molecular cytogenetic tools and their applications in ornamental plants. Plant Cell Reports 9: 1477–1488.

Zhu, J. Yu, D. & X. Xu. 2015. The phylogeographic structure of *Hydrilla verticillata* (Hydrocharitaceae) in China and its implications for the biogeographic history of this worldwide-distributed submerged macrophyte. BMC Evolutionary Biology 95:1–11.

Zwickl, D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.

Figure Legends

Figure 1. Distribution of *Najas guadalupensis* subsp. *guadalulpensis* (grey), *N. guadalupensis* subsp. *olivacea* (green) and *N.guadalupensis* subsp. *floridana* (pink) in America. Circles represent isolated occurances. Map redrawn from Lowden (1986), Urquiola Cruz (1988) and Haynes (2000) on a World Millar Cylindrical Projection using ArcMap as implemented in ArcGIS 10 (ESRI: Environmental Systems Research Institute, Redlands, California). Insert photograph is of *N. guadaluensis* subsp. *guadalupensis*.

Figure 2. Phylogenetic analysis of non-redundant combined chloroplast haplotypes (*rbcL, trnK/matK, trnL-F*) representing 197 *Najas guadalupensis* s.l. accessions in North America and Honduras, along with maps showing the ranges and number of individuals associated with each haplotype. Support values for nodes are indicated as Bayesian posterior probability (upper numbers) and maximum likelihood bootstrap (lower number) values, with hash symbols indicating poorly supported nodes. Shaded ovals represent *' floridana'* (pink) and '*olivacea'* (green). Georeferenced records (WGS_1984) are displayed on a North American Lambert Conformal Conic projection using ArcMap as implemented in ArcGIS 10 (ESRI: Environmental Systems Research Institute, Redlands, California). As numerous records overlap, Appendix A should be consulted for detailed locality information.

Figure 3. Phylogenetic analysis of non-redundant nrITS ribotypes representing 197 accessions of *Najas guadalupensis* s.l., with heatmap of ribotype distribution across chloroplast haplotype clades (upper cladogram). Support for nodes is indicated as Bayesian posterior probability (upper numbers) and maximum likelihood bootstrap (lower number) values, with stars indicating well supported nodes. Shaded ovals indicate the ribotypes and haplotypes associated with *'floridana'* (pink) and '*olivacea'* (green). Most individuals were polymorphic for the nrITS region and so haplotype counts do not equal ribotype counts.

Figure 4. A) NeighborNet equal angle network with uncorrected p-distances for Nflex-ITS repeat-specific primers sequenced in 79 *Najas. guadalupensis* s.l accessions, along with maps showing distributions of each amplicon type, and pie charts indicating haplotype distribution. In addition to amplifying *N. flexilis* and *N. canadensis* alleles, these primers amplified the *'floridana'* allele (pink) and divergent sequences (grey). Included in the network for comparison are ribotypes associated with *N. flexilis*, *N. canadensis* and *N. wrightiana* (Les et al. 2015), along with 14 ribotypes amplified with the Nguad-ITS repeat-specific and universal primers (purple) in this study. B) Map of North America showing current range of *N. canadensis* and *N. flexilis*. Both taxa are sympatric across much of their range in North America, with *N. flexilis* having a more northerly extension into Alaska. A blue triangle indicates a fossil record for *N. flexilis* in Georgia (c. 21,000 ka) (Watts 1970, Les et al. 2015). C) Amplification table for Nflex-ITS repeat-specific primers showing percentage amplification in 182 gel scored individuals, and number sequenced per haplotype. D) Representative seed images (from left to right) of *N.*

flexilis, *N. canadensis*, ‘*olivacea’*, *N. guadalupensis* subsp. *guadalupensis*, ‘ *floridana’* and N. *wrightiana*. Appendix G may be consulted for range of variation in *N. guadalupensis* s.l. seed.

Figure 5. Percentage of individuals in *Najas canadensis* (n=78), *N. flexilis* (n=49) and *N. guadalupensis* s.l. (n=177) with seed, female flowers, male flowers and nodes with adventitious roots. Boxplots above represent counts per individual within each taxon, and are proportional to sample size; with thick lines indicating the median, boxed regions showing the interquartile ranges, top whiskers (where present) indicating 1.5 times the interquartile range and outliers as black circles. Jitterplots with a scatter of 0.2 illustrate individual counts within each taxon. Counts represent the first 12 nodes of a single mature shoot from herbarium specimens. See Appendices A and J for accession numbers.

Figure 6. Percentage of *Najas* canadensis (n=78), *N. flexilis* (n=49) and *N. guadalupensis* s.l. accessions (n=177) with seed, female flowers, male flowers and adventitious roots on the first 12 nodes of a single mature shoot. Node one refers to the uppermost node, with node 12 indicating the lowest node.

Figure 7. Percentage of individuals within 16 *N. guadalupensis* s.l. chloroplast haplotypes with seed, male flowers and female flowers . Numbers beside haplotype indicates sample size.

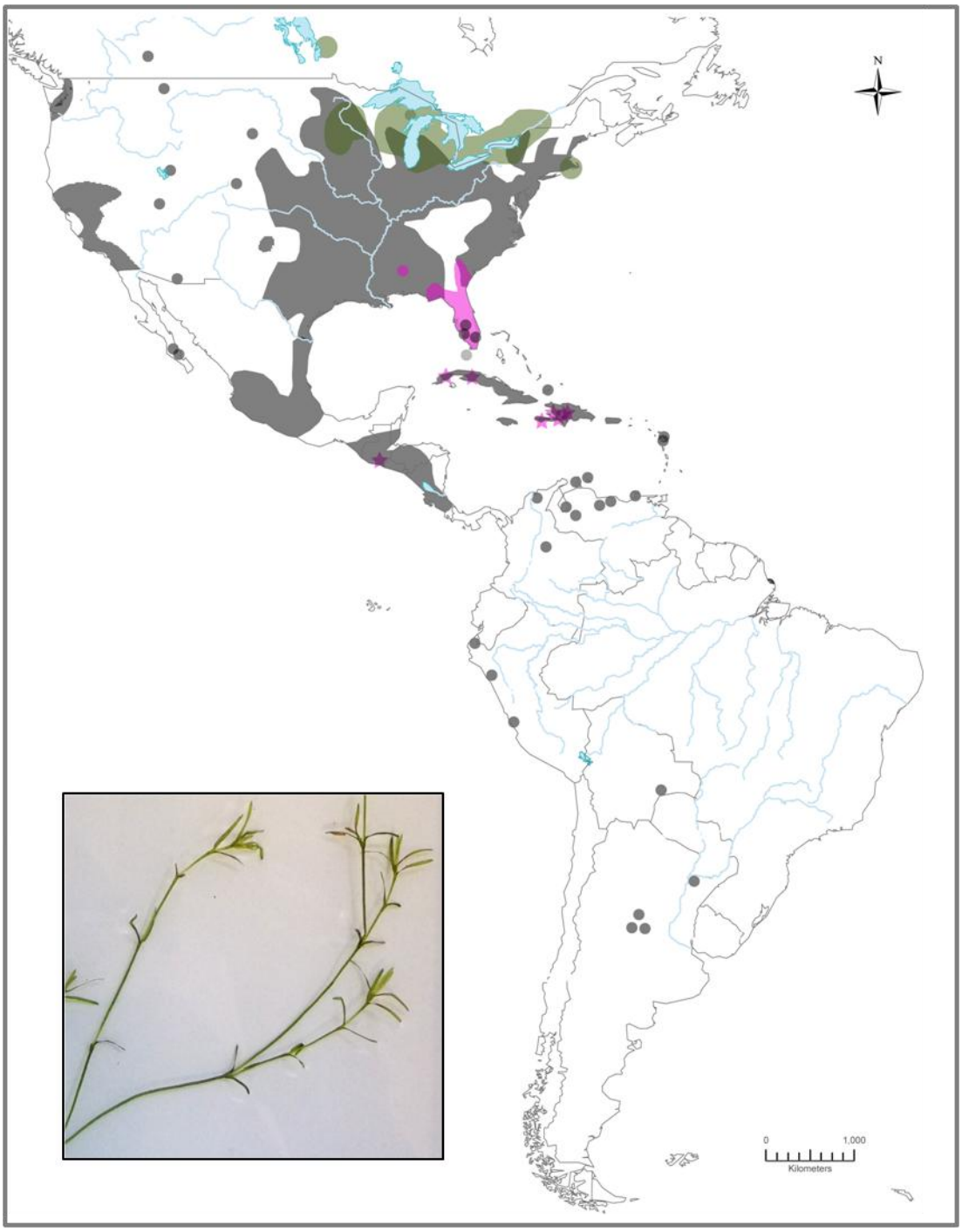


Figure 1

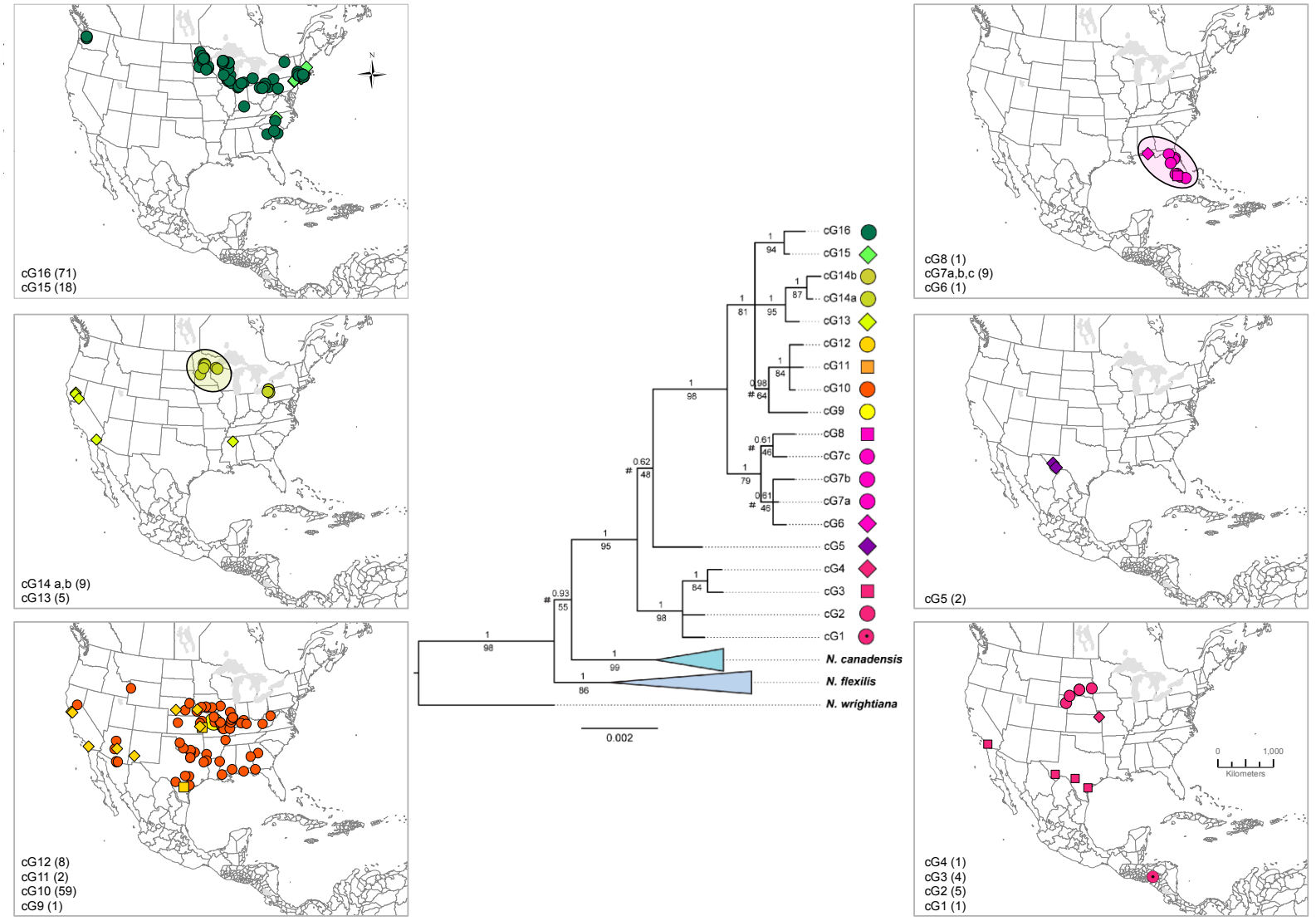

cG16 (71)
cG15 (18)
cG14 a,b (9)
cG13 (5)
cG12 (8)
cG11 (2)
cG10 (59)
cG9 (1)
cG8 (1)
cG7a,b,c (9)
cG6 (1)
cG5 (2)
cG4 (1)
cG3 (4)
cG2 (5)
cG1 (1)
cG16
cG15
cG14b
cG14a
cG13
cG12
cG11
cG10
cG9
cG8
cG7c
cG7b
cG7a
cG6
cG5
cG4
cG3
cG2
cG1
N. canadensis
N. flexilis
N. wrightiana
0.002
0
1,000
Kilometers

Figure 2

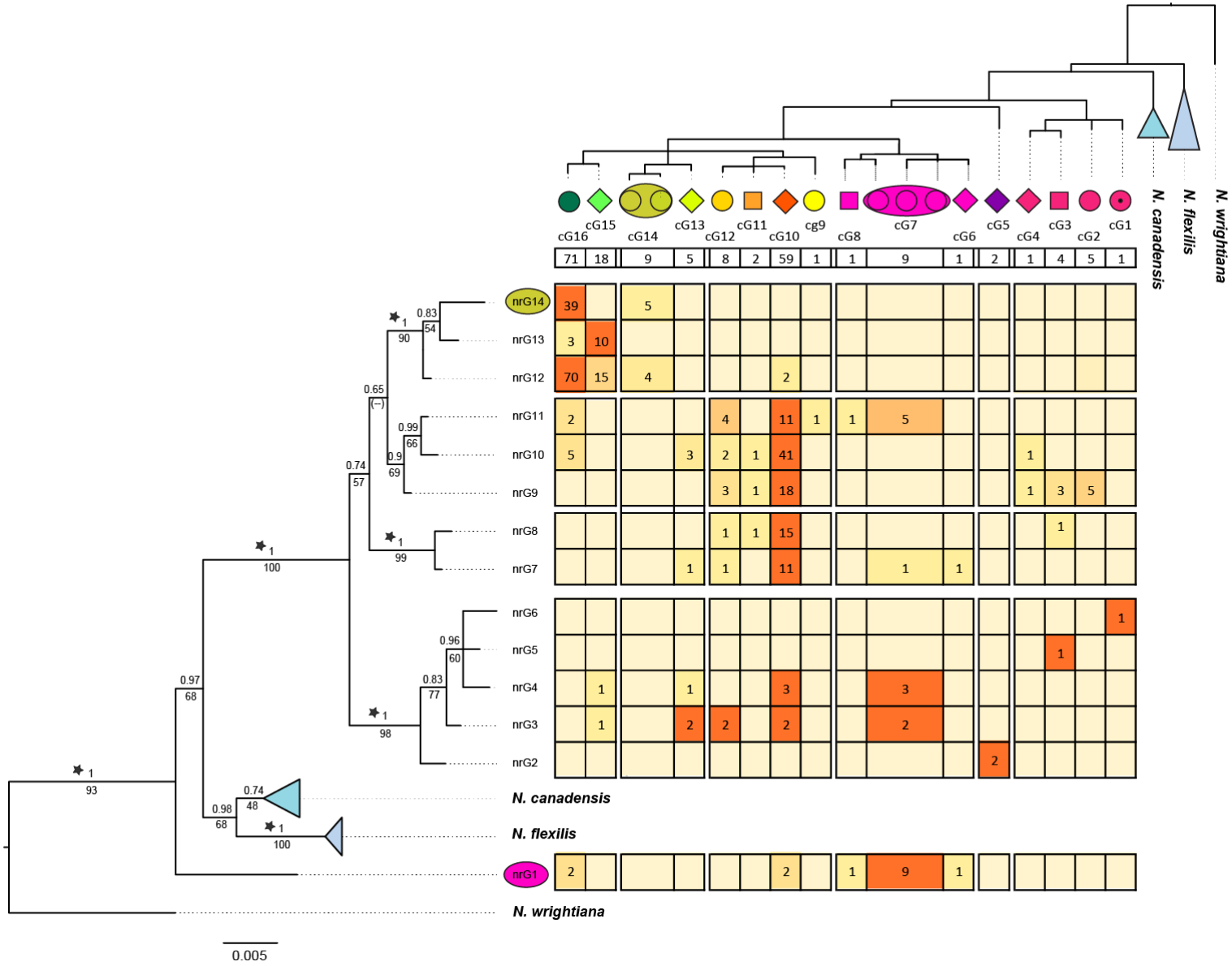
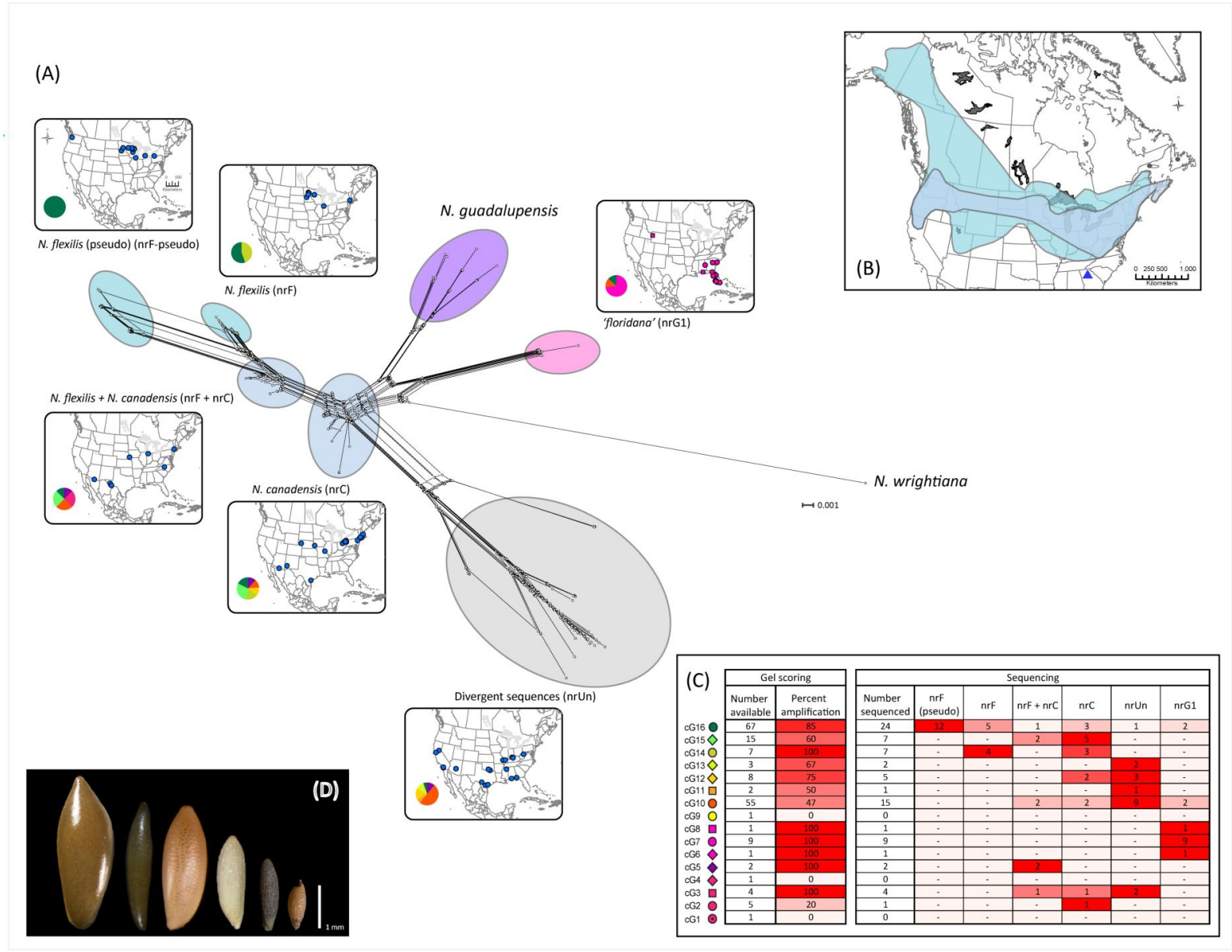

cG16
cG15
cG14
cG13
cG12
cG11
cG10
cg9
cG8
cG7
cG6
cG5
cG4
cG3
cG2
cG1
N. canadensis
N. flexilis
N. wrightiana
71 18 9 5 8 2 59 1 1 9 1 2 1 4 5 1
nrG14 39 5
nrG13 3 10
nrG12 70 15 4 2
nrG11 2 4 11 1 1 5
nrG10 5 3 2 1 41 1
nrG9 3 1 18 1 3 5
nrG8 1 1 15 1
nrG7 1 1 11 1 1
nrG6 1
nrG5 1
nrG4 1 1 3 3
nrG3 1 2 2 2 2
nrG2 2
N. canadensis
N. flexilis
nrG1 2 2 1 9 1
N. wrightiana
0.005

Figure 3



	Gel scoring		Sequencing						
	Number available	Percent amplification	Number sequenced	nrF (pseudo)	nrF	nrF + nrC	nrC	nrUn	nrG1
cG16	67	85	24	12	5	1	3	1	2
cG15	15	60	7	-	-	2	5	-	-
cG14	7	100	7	-	4	-	3	-	-
cG13	3	67	2	-	-	-	-	2	-
cG12	8	75	5	-	-	-	2	3	-
cG11	2	50	1	-	-	-	-	1	-
cG10	55	47	15	-	-	2	2	9	2
cG9	1	0	0	-	-	-	-	-	-
cG8	1	100	1	-	-	-	-	-	1
cG7	9	100	9	-	-	-	-	-	9
cG6	1	100	1	-	-	-	-	-	1
cG5	2	100	2	-	-	2	-	-	-
cG4	1	0	0	-	-	-	-	-	-
cG3	4	100	4	-	-	1	1	2	-
cG2	5	20	1	-	-	-	1	-	-
cG1	1	0	0	-	-	-	-	-	-

Figure 4

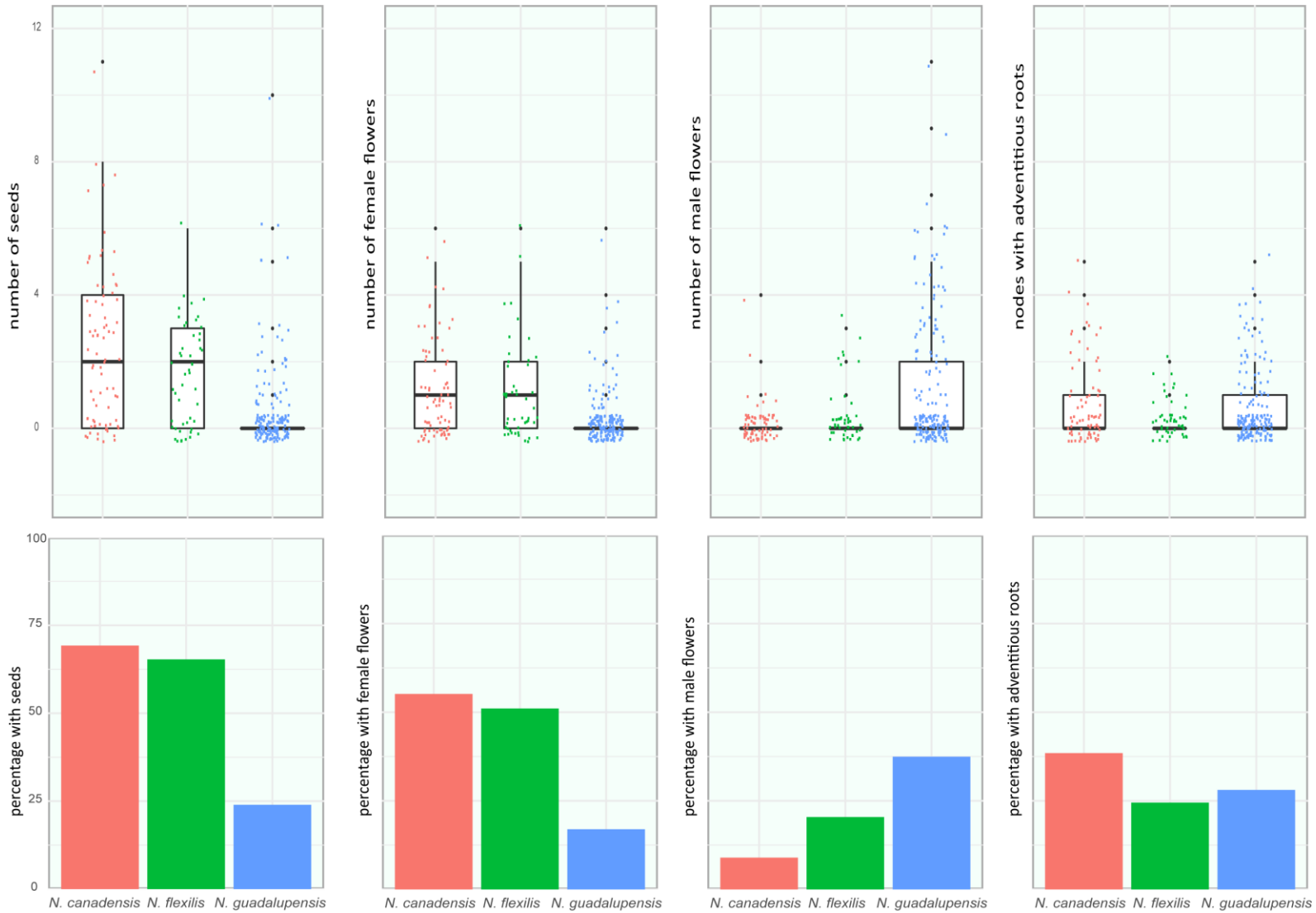

number of seeds
number of female flowers
number of male flowers
nodes with adventitious roots
12
8
4
0
percentage with seeds
percentage with female flowers
percentage with male flowers
percentage with adventitious roots
100
75
50
25
0
N. canadensis
N. flexilis
N. guadalupensis

Figure 5

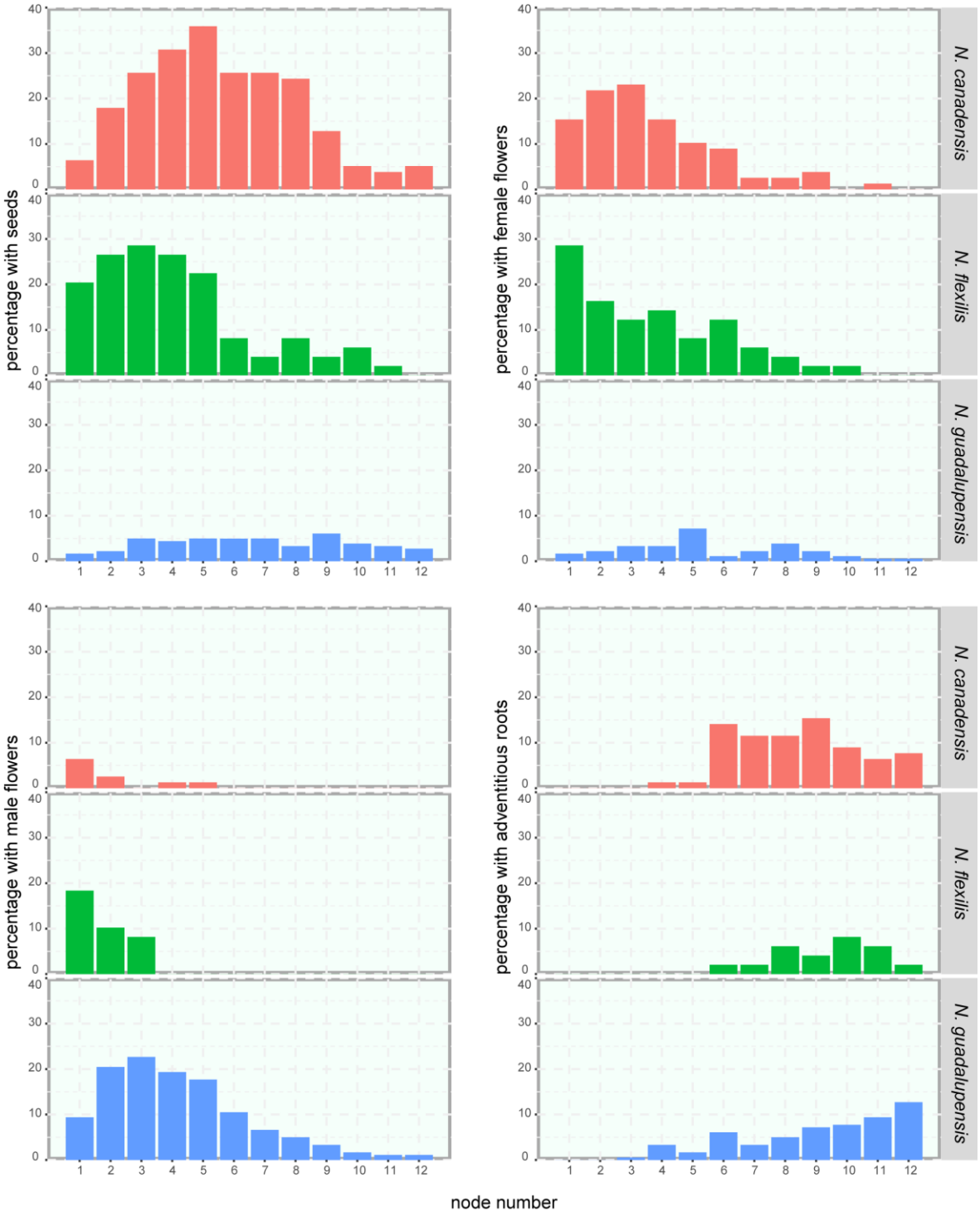

percentage with seeds
percentage with female flowers
percentage with male flowers
percentage with adventitious roots
N. canadensis
N. flexilis
N. guadalupensis
node number

Figure 6

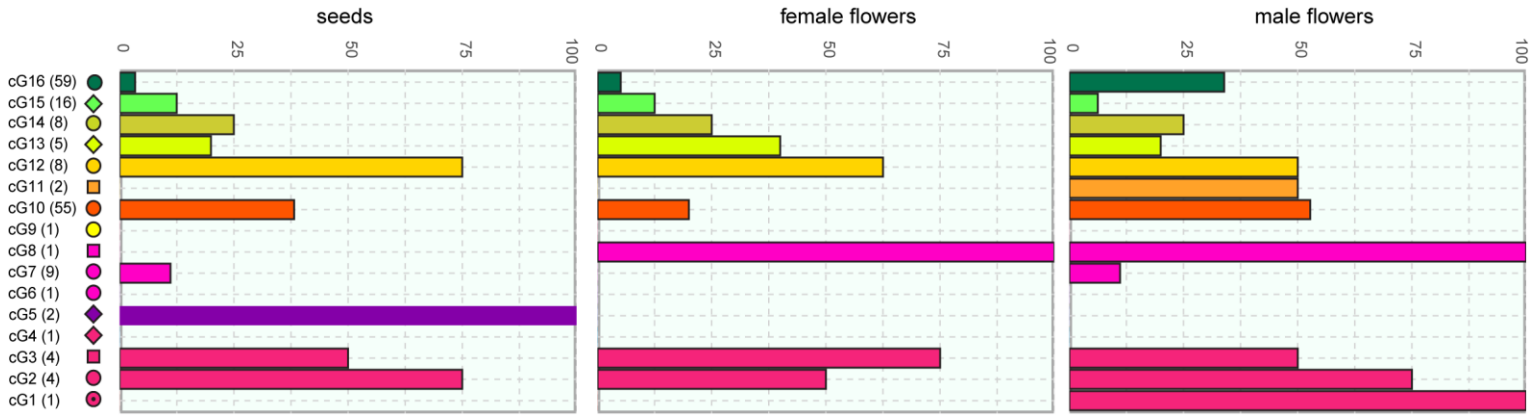

seeds
female flowers
male flowers
0
25
50
75
100
cG16 (59)
cG15 (16)
cG14 (8)
cG13 (5)
cG12 (8)
cG11 (2)
cG10 (55)
cG9 (1)
cG8 (1)
cG7 (9)
cG6 (1)
cG5 (2)
cG4 (1)
cG3 (4)
cG2 (4)
cG1 (1)

Figure 7

Adaptive evolution of the chloroplast genome in the submersed monocotyledon *Najas* (Hydrocharitaceae)

Ursula M. King[1, 3], Donald H. Les[1], Elena L. Peredo[2], and Lori K. Benoit[1]

[1]Department of Ecology and Evolutionary Biology, University of Connecticut, 75 North Eagleville Rd., Storrs, CT 06269-3043, U.S.A.; email: les@uconn.edu
[2]Marine Biological Laboratory, 7 Marine Biological Laboratory St., Woods Hole, MA 02543, U.S.A.; email: elperedo@hotmail.com
[3]Author for correspondence; email: ursula.king@uconn.edu

Abstract. *Najas* (Hydrocharitaceae), a cosmopolitan genus of about 40 species, stands among the most intricately adapted of aquatic angiosperms. All species are dioecious or monoecious and are entirely submersed and pollinated underwater (a form of hydrophily). Unlike most Hydrocharitaceae, *Najas* species are annuals that rely primarily on seed for reproduction. Most of their anatomical and morphological features are characterized by extreme reduction. Submersed plants like *Najas* face unique physiological challenges as they grow across a broad ecological gradient ranging from shallow, warm, and bright waters to deep, dark, and cold waters. How these plants maintain essential metabolic processes under such diverse conditions has not yet been explained in any detail. We are beginning to investigate this question by comparing sequence data obtained from complete chloroplast genomes of *Najas* and related taxa. Already this work has provided evidence of unique alterations of the chloroplast *psaA*/*psaB* operon and the loss/pseudogenization of photosynthetic genes that otherwise are conserved strongly across terrestrial plants. In addition to reviewing the adaptive implications of these features, we report new evidence to demonstrate that plastid coding regions of *Najas* species diverged by positive selection with respect to those of other Hydrocharitaceae, monocotyledons, and angiosperms. Ultimately, the objective of this work is to identify key regions of the chloroplast and nuclear genomes that have facilitated the major ecological transition of flowering plant species from life on land to life in the water.

Keywords: aquatic angiosperms, cpDNA genome, d_N/d_S, plastid-encoded polymerase, *rpoB*, *rpoC*$_1$

Introduction

The evolutionary history of land plants largely chronicles their terrestrialization. As they diverged from their aquatic algal ancestors some half-billion years ago, land plants continually acquired essential adaptations that facilitated their acclimation to terrestrial habitats (Delaux et al., 2012). Angiosperms are a pinnacle of this adaptive trajectory. In addition to a cuticle and stomata, so vital for early land plant survival (Delaux et al., 2012), they also gained vascular tissues, seeds, pollen, and nonmotile sperm which were invaluable innovations for enduring a desiccating terrestrial environment (Evert & Eichhorn, 2012). However, in what might be termed evolutionary irony, a small number of angiosperm lineages (fewer than two percent of species) have successfully recolonized aquatic habitats in repeated, secondary radiations (Cook, 1996, 1999). The recolonization of aquatic habitats represents a remarkable feat, given that it was necessary for water plants to overcome an extensive background of attributes honed for a terrestrial existence over millions of years of adaptive evolution. How did this reverse transition come about? The extremely low number of aquatic angiosperm species indicates that it was not achieved without great difficulty.

Morphological adaptation is the most conspicuous feature of aquatic angiosperms, and many water plants exhibit extensive phenotypic plasticity (Sculthorpe, 1967). Some species are able to produce such highly differentiated phenotypes that they have merited taxonomic distinction. *Gratiola aurea* Pursh (Plantaginaceae) typically grows on exposed wet shores, where it produces erect, aerial stems with essentially unmodified insect-pollinated flowers; however, it also can persist in deep water as entirely submersed, dwarf, sterile plants known as *G. aurea* forma *pusilla* Fassett (Fassett, 1957). A similar extent of plasticity also characterizes *Hypericum boreale* (Britton) E.P. Bicknell (Hypericaceae) and its sterile, submersed counterpart *H. boreale* forma *callitrichoides* Fassett, as well as *Pontederia cordata* L. (Pontederiaceae), with its submersed variant *P. cordata* forma *taenia* Fassett (Fassett, 1957). Exposed shoreline plants of *Glossostigma cleistanthum* W.R. Barker (Phrymaceae)

http://dx.doi.org/10.21135/893275341.004

produce long-stalked, chasmogamous flowers and short leaves, while submersed plants produce significantly longer leaves and sessile, cleistogamous flowers (Les et al., 2006). By their ability to transform morphologically through phenotypic plasticity, these examples indicate how some aquatic species survive under the substantially diverse conditions that differentiate terrestrial and underwater habitats.

Other aquatic plants have achieved comparable success through the evolution of heterophylly, an adaptation that enables a single individual to discretely alter its leaf morphology and anatomy simultaneously with respect to different ambient habitat conditions (Sculthorpe, 1967). Heterophyllous aquatic species typically produce highly dissected foliage when growing underwater but develop broad, floating leaves as their foliage contacts an aerial or surface environment. Because each leaf type functions optimally under the specific environmental conditions that induce its morphology (Sculthorpe, 1967; Hutchinson, 1975), a single plant is able to exploit different points along the aquatic environmental spectrum more effectively. Heterophylly also is associated with physiologically related biochemical features. The acquisition of secondary metabolites such as glycoflavones is believed to have been instrumental in the successful colonization of land by plants (Swain, 1970, 1975). Like many terrestrial plants, the floating leaves of heterophyllous pondweed (*Potamogeton*) species are exposed to intense light and contain glycoflavones, which safeguard plants against harmful effects of ultraviolet radiation (Les & Sheridan, 1990). However, their underwater leaves often are devoid of glycoflavones, presumably because efficient absorption of ultraviolet light by water has relaxed selection to retain the compounds in submersed foliage (Les & Sheridan, 1990).

Yet, many aquatic plants thrive across a gradient of depths ranging from only a few centimeters to nearly 5 m (Nichols, 1999) without being heterophyllous or otherwise exhibiting strong phenotypic differentiation. One such example is *Najas*, a genus of roughly 40 species, which possesses a highly specialized anatomy and morphology modified to accommodate a life history that occurs entirely under water (Les et al., 2010). *Najas* species are highly simplified structurally, to the extent of virtually lacking xylem tissue or lignification and possessing small, linear leaves reduced to only two to three cell layers (Sculthorpe, 1967; Ogden, 1974). Their annual habit is associated with a lack of vegetative propagules, which are common in other submersed plants (Sculthorpe, 1967). *Najas* species also are water pollinated (hydrophilous), a reproductive mechanism unique to hydrophytes, but found only in fewer than five percent of all aquatic angiosperms (Les, 1988; Les et al., 2010). Consequently, although they were derived from terrestrial ancestors, the specialized, reduced *Najas* species have been transformed into plants that survive only briefly if removed from the water (Sculthorpe, 1967). Nevertheless, the genus clearly excels under water as evidenced by its relatively high species diversity (Les, 1988) and broad environmental tolerances (Nichols, 1999).

With respect to terrestrial species, submersed hydrophytes must overcome such obvious constraints as sediment anoxia, carbon limitation, light attenuation, and increased hydrostatic pressure (Hutchinson, 1975). Reduced (or absent) xylem and root systems and complete lack of functional stomata combine to preclude an operational transpiration stream in submersed aquatics (Sculthorpe, 1967). Hence, it is apparent that in addition to morphological specialization, the physiology of aquatic angiosperms also has adapted to enhance survival under conditions never encountered by the majority of plant species, which are terrestrial.

Submersed plants represent an ideal system for studying photosynthetic adaptation due to the unusual selective pressures imposed by aquatic habitats. Many species colonize a broad depth range where conditions can vary significantly from bright and warm, shallow sites to cold and dark, deep waters (Fig. 1A). Different organs of even a single aquatic plant can experience highly contrasting conditions simultaneously, throughout the vertical water column (Fig. 1B). At the height of day, leaves growing near the water surface encounter more intense temperatures and light levels due to the absence of a shading vegetative canopy. However, leaves along lower regions of the shoot grow under light levels and water temperatures that decrease continuously with depth (Vercauteren et al., 2011). Similarly, individuals of a single submersed species might occupy different depths, or might occur in shallow or deep waters in different years (Fig. 1A).

Nonetheless, aquatic plants somehow tolerate variable water depths remarkably well and often flourish under quite different light and temperature regimes. The environmental latitude of conditions in aquatic habitats presents an apparent adaptive paradox to submersed plants; that is, how can both sun-adapted (shallow water) and shade-adapted (deep water) characteristics evolve simultaneously? How does one genome enable a species to carry out photosynthesis under extremely high light and warm temperatures of shallow waters but also under cold, near dark conditions at greater depths? Do aquatic plants acquire such broad habitat tolerances by evolving new traits, by

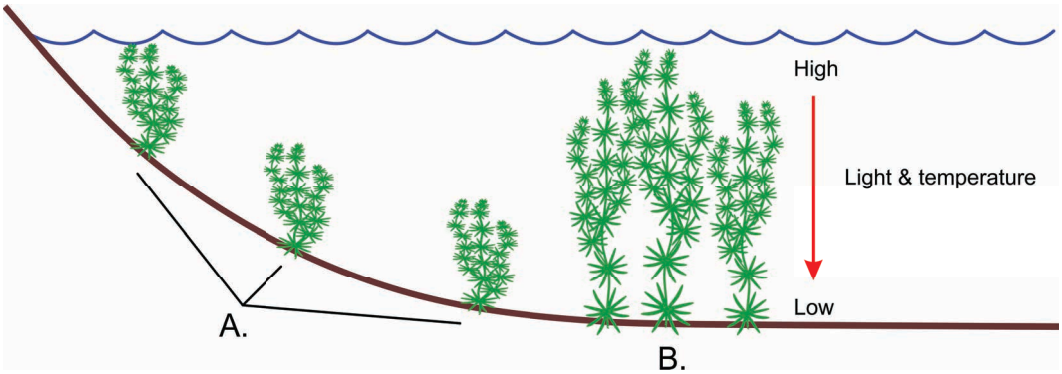


FIGURE 1. Submersed aquatic plants exist under extreme and variable environmental conditions. **A.** A single species can maintain photosynthesis in shallow (high light and temperature) or deep (low light and temperature) waters. **B.** In species with elongate shoots, various portions of individual plants can be exposed to different extremes simultaneously.

expressing multiple genes, or by altering single enzymes to function optimally across the entire light and temperature gradient? These are questions that we have only begun to investigate.

Photosynthesis arguably is the most important physiological process carried out by autotrophic plants. Shifts to parasitism for example, have resulted in substantial genomic modifications, especially in the plastid, where essential photosynthetic genes normally reside (DePamphilis & Palmer, 1990). Because of the profound influence of water depth on light and temperature, it is reasonable to think that physiological adaptations in higher aquatic plants might be elucidated by focusing on the chloroplast (cp) genome, where critical photosynthetic and related "housekeeping" genes are retained from ancestral cyanobacteria (De Las Rivas et al., 2002; Peredo et al., 2012).

Such adaptations could be manifested in various ways through gene expression. Physiological adaptations could be achieved by the acquisition of novel traits, such as innovative mechanisms capable of regulating gene expression in response to different environmental stimuli. Akin to photosynthetic gene loss in parasitic plants, which no longer are under selection to maintain an autotrophic physiology (DePamphilis & Palmer, 1990), aquatic plants also would be expected to lose or trivialize the function of genes that originated primarily as terrestrial adaptations, either by their dissolution or through diminished expression.

Furthermore, it is reasonable to predict that the structures of photosynthetic-related enzymes in aquatic plants might diverge substantially from those optimized to function in a terrestrial environment, due to the need for these catalysts to maintain efficiency across a broader range of temperature and the light conditions potentially available during each growing season (Fig. 1A). The ability of aquatic plants to maintain essential processes like transcription while growing at different temperatures would implicate associated loci like RNA polymerase genes as likely targets of natural selection. Indeed, comparisons of chloroplast regions in *Najas* have indicated that at least some such regions (e.g., the *rpoB*/C_1/C_2 operon) appear to have diverged considerably from those of terrestrial angiosperms (Peredo et al., 2013).

Here we review recent efforts to characterize photosynthetic adaptations in submersed aquatic plants with a focus on the chloroplast genome of the highly specialized genus *Najas*. Examples relating to novel regulatory mechanisms and gene loss are presented along with new data providing evidence that individual cp genes in *Najas* have experienced strong positive selection and have undergone considerable divergence as a consequence. By offering additional insights regarding the influence of genomic selection, these examples should help to further elucidate mechanisms of plant molecular evolution. Such information should also facilitate phylogenetic research and other applications, which rely on an accurate understanding of processes that influence the divergence of genes.

MATERIALS AND METHODS

Each of four plastid-encoded RNA polymerase (PEP) subunits was screened for evidence of positive selection by computing and comparing values of omega (ω), the ratio of nonsynonymous to synonymous substitution rates expressed as d_N/d_S. Branches were divided a priori into foreground branches (branches where individual codon sites are tested for positive selection), whereas all other branches on the

tree represented the background branches. By using a likelihood ratio test (LRT), a null model constraining positive selection on the foreground branch was compared with the alternative model, in which positive selection is allowed to occur. Significance was assessed using a chi-square test.

Several different analyses were performed using PEP subunit sequence data obtained for a variety of angiosperm species (GenBank accession numbers are provided in brackets for newly and previously sequenced material). Two genera of submersed, aquatic Hydrocharitaceae (*Elodea* and *Najas*) comprised the foreground lineages (singly or in combination), which were compared against a background set comprising 31 terrestrial plant species selected from among the complete chloroplast genome sequences available in GenBank (Benson et al., 2005).

Our selection of background taxa represented the major angiosperm clades (APG, 2009) well by including: ANITA-grade groups (*Chloranthus* [EF380352], *Illicium* [EF380354]), magnoliids (*Magnolia* [JN867579], *Piper* [DQ887677]), Asparagales (*Cymbidium* [KC876123]), Dioscoreales (*Dioscorea* [EF380353]), commelinids (*Arundinaria* [JX235347], *Lolium* [JX871942], *Zingiber* [JX088661]), noncore eudicots (*Ranunculus* [DQ359689]), core eudicots (*Trochodendron* [KC608753]), fabid rosids (*Castanea* [HQ336406], *Fragaria* [JN884817], *Pyrus* [AP012207], *Vigna* [AP012598]), malvid rosids (*Arabidopsis* [AP000423], *Eucalyptus* [HM347959], *Gossypium* [JF317355], *Lepidium* [AP009374], *Oenothera* [EU262889]), noncore asterids (*Fagopyrum* [EU254477]), lamiid asterids (*Ardisia* [KC465962], *Asclepias* [JN710465], *Coffea* [EF044213], *Nicotiana* [Z00044], *Salvia* [JX312195], *Tectona* [HF567869]), and campanulid asterids (*Anthriscus* [GU456628], *Artemisia* [JX293720], *Daucus* [DQ898156], *Jacobaea* [HQ234669]).

Our foreground groups included the previously sequenced *Elodea canadensis* Michx. [JQ310743] and *Najas canadensis* Michx. [JX978472, in GenBank as *N. flexilis* (Willd.) Rostk. & W.L.E. Schmidt], as well as five additional *Najas* species: *N. flexilis, N. filifolia* R.R. Haynes, *N. gracillima* Morong, *N. marina* L., and *N. minor* All., which we newly sequenced (Table 1). This selection of *Najas* species represented all major clades within the genus (Les et al., 2010).

We summarized the relative phylogenetic positions of included taxa by extracting a tree file directly from National Center for Biotechnology Information (NCBI) taxonomy (https://ncbi.nlm.nih.gov/taxonomy), which we then edited to comply with PAML version 4.7 (Yang, 2007) program formatting. The resulting unrooted network specified by this file is illustrated in Fig. 2. Sequences for all four PEP subunits (*rpoA*, *rpoB*, $rpoC_1$, and $rpoC_2$) were aligned by their inferred amino acid composition (*rpoA* = 1005 nts; *rpoB* = 3291 nts; $rpoC_1$ = 2094 nts [exon 1 and exon 2]; $rpoC_2$ = 4884 nts) using Geneious V6.1.6 (Biomatters Limited, http://geneious.com; Drummond et al., 2010).

All analyses for positive selection were conducted using the program PAML 4.7 (Yang, 2007) under the branch-site model (test 2 in codeml) with the following program settings: model = 2, NSsites = 2; null hypothesis: fix_omega = 1, omega = 1; alternative hypothesis: fix_omega = 0 (estimated), omega = 1.5. An LRT (twice the difference of the log likelihood of compared models; $2\Delta\mathbf{ln}$, $df = 1$) was conducted. Significance of results was determined from a chi-square distribution using a critical value of 3.84 with 0.05 as the threshold P significance value. The Bayes empirical Bayes (BEB) method (Yang et al., 2005) was used to calculate posterior probabilities for site classes, whenever the LRT suggested the presence of codons under positive selection on the foreground branch.

TABLE 1. Accessions and GenBank numbers of *Najas* L. used in molecular analyses.

Taxon	Voucher	*rpoA*	*rpoB*	*rpoC1*	*rpoC2*
Najas marina L.	*Zdenek Kaplan* 10/316 (CONN)	KM373917	KM373902	KM373907	KM373912
Najas filifolia R.R. Haynes	*D. Les* 756 (CONN)	KM373918	KM373903	KM373908	KM373913
Najas flexilis Rostk. & W.L.E. Schmidt	*R.K. Shannon* 1157 (CONN)	KM373919	KM373904	KM373909	KM373914
Najas minor All.	*R.K. Shannon* 1251 (CONN)	KM373920	KM373905	KM373910	KM373915
Najas gracillima (A. Braun ex Engelm.) Magnus	*D. Les* 931 (CONN)	KM373921	KM373906	KM373911	KM373916

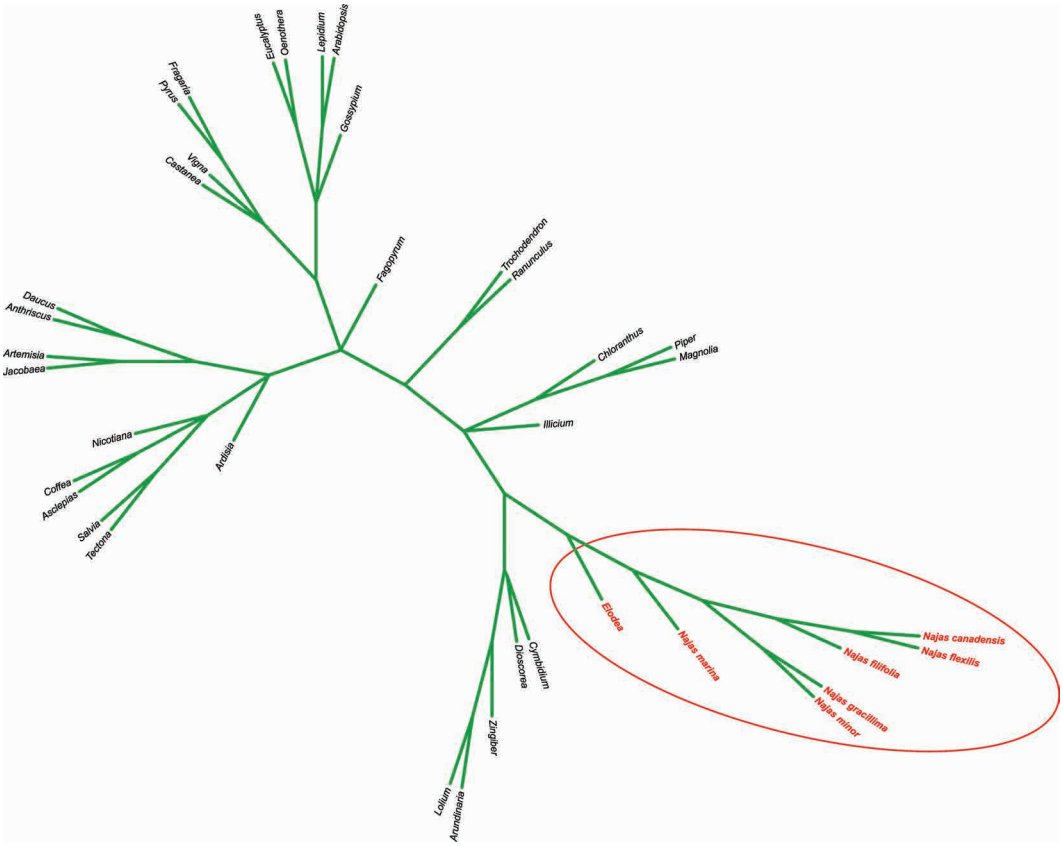


FIGURE 2. Unrooted network used to analyze d_N/d_S ratios of plastid-encoded RNA polymerase subunits in a taxonomically diverse assemblage of terrestrial and aquatic angiosperms. The two aquatic genera (*Najas* and *Elodea*) occur in sister clades of the family Hydrocharitaceae (circled) and were used individually or in combination as foreground taxa (see Table 2); the remaining terrestrial representatives were assigned to the background.

Twenty individual analyses were conducted, representing five different comparisons made for each of the four plastid *rpo* subunits: (1) a foreground branch comprising *Elodea* and six *Najas* species with 31 terrestrial background taxa; (2) *Elodea* as the foreground branch with six *Najas* species included with 31 terrestrial taxa in the background; (3) six *Najas* species as the foreground branch with *Elodea* included with 31 terrestrial taxa in the background; (4) *Elodea* as the foreground branch with 31 terrestrial taxa (all *Najas* species excluded) in the background; (5) six *Najas* species as the foreground branch with 31 terrestrial taxa (*Elodea* excluded) in the background.

RESULTS

Evidence of positive selection (Table 2) was detected in the PEP subunit *rpoB* for two scenarios (see above): comparison 3, in which the *Najas* species were designated as the foreground branch and *Elodea* along with 31 terrestrial taxa as the background; and comparison 5, in which the *Najas* species comprised the foreground branch and 31 terrestrial taxa (*Elodea* excluded) represented the background. The BEB method indicated positive selection at 12 *rpoB* codon positions (1% of total) in comparison 3 and at 10 codon positions in comparison 5 (Fig. 3). In comparison 3, five of the nonsynonymous substitutions (45% of significant sites) resulted in a change of charge in amino acids of *Najas* relative to all other taxa; two additional sites reflected a change of charge in *Najas* relative to a subset of the background genera (Fig. 3). A highly divergent region occurred from codon alignment positions 197–207, where six closely situated substitutions clustered, including four effecting a change in charge (Fig. 3).

Evidence of positive selection also was obtained for the *rpoC1* subunit in the same two scenarios (comparisons 3 and 5), with the BEB analysis indicating selection at 13 sites (2% of total) in both analyses.

TABLE 2. Comparative ω values from different branch-site tests for positive selection on ribosomal polymerase (*rpo*) subunits.

Foreground vs. background					
Locus	*Elodea* + *Najas* (6 spp.) vs. 31 taxa	*Elodea* vs. 37 taxa (incl. 6 *Najas* spp.)	*Najas* (6 spp.) vs. 38 taxa (incl. *Elodea*)	*Elodea* vs. 31 taxa (excl. *Najas*)	*Najas* (6 spp.) vs. 31 taxa (excl. *Elodea*)
rpoA	1.0 (NS)	1.515 (NS)	1.0 (NS)	2.874 (NS)	1.0 (NS)
rpoB	1.0 (NS)	2.084 (NS)	**3.770*** **	1.0 (NS)	**2.862***
rpoC$_1$	1.0 (NS)	1.0 (NS)	**3.628*** **	1.0 (NS)	**2.494***
rpoC$_2$	1.0 (NS)	– (–)	1.0 (NS)	1.0 (NS)	1.0 (NS)

The P values from the corresponding likelihood ratio tests are indicated as: (NS) = > 0.05; * = ≤ 0.05; ** = < 0.01; *** = ≤ 0.001. Significant values are highlighted in bold. – = ω undefined.

Eight of these substitutions (62% of significant sites) altered the charge relative to all background lineages including *Elodea*; four additional sites reflected a change of charge in *Najas* relative to a subset of the background genera (Fig. 3). A highly divergent region (codon alignment positions 224–231) also was detected in *rpoC*$_1$, where six significant nonsynonymous substitutions occurred, with four effecting a change in charge (Fig. 3). No evidence of positive selection could be detected for any of the PEP subunits when *Elodea* was included in the foreground, nor in any comparison involving subunits *rpoA* or *rpoC*$_2$ (Table 2).

DISCUSSION

Although the breadth of knowledge in plant physiology has grown extensively over the past century, there has been little elucidation of photosynthetic adaptations specific to aquatic plants. It is predictable that major adaptive changes should be evident in the photosynthetic pathway of submersed aquatic plants because they occupy habitats where light and carbon availability differ considerably from terrestrial conditions.

However, the elucidation of photosynthetic-related adaptations in water plants represents a complex problem, due to the ability of various species to incorporate C_3, C_4, and Crassulacean acid metabolism pathways as well as bicarbonate uptake (Pedersen et al., 2013). It is well known that some submersed plants use the bicarbonate ion as a carbon source associated with its increased prevalence in increasingly alkaline waters (Pedersen et al., 2013). Yet even here the mechanism of bicarbonate uptake varies, with some species capable of direct bicarbonate uptake but others first requiring the conversion to carbon dioxide at the leaf diffuse boundary layer (Pedersen et al., 2013).

More subtle adaptations to aquatic photosynthesis are likely to be difficult to detect and characterize. In addition to the 100 or so photosynthesis-related genes coded by the cp genome (McFadden, 2001), as many as 5000 nuclear genes also carry out some type of plastid-related function (Kuroda & Maliga, 2003; Bock & Timmis, 2008). Consequently, such a large network of potentially interacting genes should provide a broad selective arena within which to evolutionarily modify and hone the photosynthetic processes of water plants. Our objective here is to stimulate interest in these processes by reviewing examples of putatively adaptive modifications specific to the cp genome itself, a subject not discussed previously in any detail. Hopefully these examples of novel functions, adaptive loss of function, and modification through positive selection will further advance the understanding of physiological adaptations in aquatic plants along with associated theoretical implications.

THE *PSAA*/*PSAB* SPACER IN *NAJAS*: A MODIFIED TRANSLATIONAL APPARATUS

The adaptive significance of novel, environmentally responsive traits has been recognized since the time of Darwin (West-Eberhard, 2008). However, complex genetic interactions can make it difficult to precisely elucidate specific genetic changes associated with novel function (Filatov et al., 2006). Although the search for novel, environmentally responsive adaptations in submersed aquatic plants remains in early stages of inquiry, a unique molecular feature has been identified in *Najas*, which constitutes a simple and straightforward genetic basis, yet potentially represents an important photosynthetic adaptation in these submersed angiosperms (Peredo et al., 2012).

Photosynthesis in higher plants relies inexorably on the assembly of photosystem I (PSI), which mediates linear and cyclic electron transfer pathways

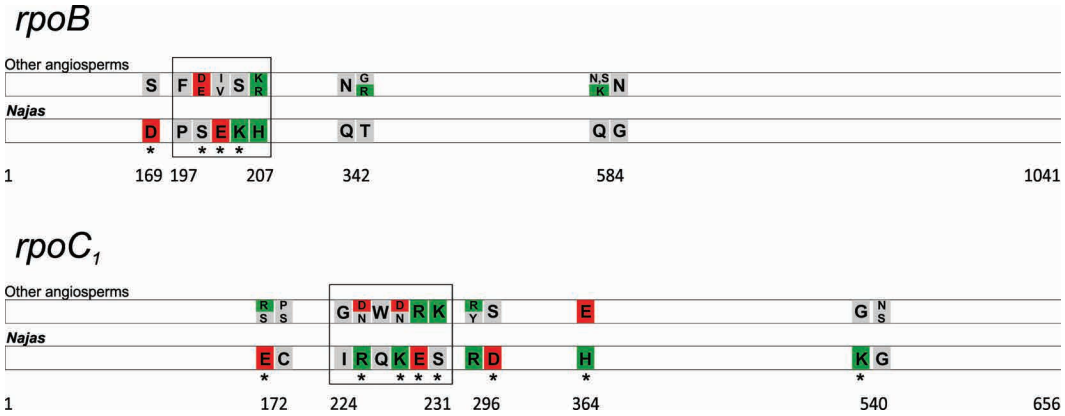


FIGURE 3. In *Najas*, two of the plastid-encoded RNA polymerase subunit genes have diverged considerably under positive selection (boxed region) as indicated by significant ω scores relative to other angiosperms (boxed region; see text). Each schematic indicates approximate relative alignment positions (*N*-terminal end at left; gaps removed) of all significantly divergent codons ($P < 0.01–0.05$) with the corresponding amino acids (AA) abbreviated by standard International Union of Pure and Applied Chemistry codes; charges are indicated by color: gray = neutral; pink = negative; green = positive. Asterisks denote substitutions in *Najas* that represent an altered charge relative to all other taxa surveyed. Both *rpoB* (1041 AA screened) and *rpoC1* (656 AA screened) show highly divergent regions (boxed) where numerous nonsynonymous and charge-altering amino acid substitutions are concentrated. Despite several synonymous nucleotide substitutions, the six *Najas* species examined (representing all major clades within the genus) maintained identical amino acids at the sites indicated.

(Shikanai, 2014). Accordingly, PSI has been characterized as a critical component in the regulation of photosynthesis (Shikanai, 2014). In angiosperms, PSI is a complex involving numerous structural protein subunits and antennal light harvesting proteins (Kouřil et al., 2014). The critical reaction core of PSI comprises two large protein subunits encoded by the chloroplast genes *psaA* and *psaB* (Semenov, 2012). Formation of the *psaA*/*psaB* heterodimer initiates the assembly of PSI and subsequently directs the attachment of other subunits to the complex (Schwabe & Kruip, 2000; Marín-Navarro et al., 2007). For PSI assembly to proceed normally, an equal stoichiometric ratio of *psaA* to *psaB* must be maintained. Otherwise, the reduced expression of *psaB* concomitantly represses translation of *psaA* by an autoregulatory mechanism (Choquet & Wollman, 2002; Wostrikoff et al., 2004; Marín-Navarro et al., 2007; Midorikawa et al., 2009). Consequently, it is clear that the regulation of *psaA*/*psaB* expression would provide an effective means of controlling PSI assembly and thus overall photosynthetic rates. One such mechanism (summarized below) is based on results reported recently by Peredo et al. (2012) for the submersed genus *Najas*.

In plants, *psaA* and *psaB* are cotranscribed with *rps14* in a tricistronic operon (Meng et al., 1988). Transcription of these genes occurs by means of a PEP, which yields a single mRNA transcript that is unmodified post-transcriptionally. An extreme level of conservation, the possession of a strong ribosomal binding site, and a cis-regulatory mRNA folding pattern have ascribed a regulatory function to the *psaA*/*psaB* intercistronic spacer in translation of the *psaB* subunit. All *Najas* species investigated to date possess an unusual three nucleotide insert in the mid-region of the spacer, which has not been found in any other plant group investigated. In silico modeling of the mRNA transcript of *Najas* results in different folding configurations of the spacer transcript at different temperatures. At high temperatures (which approximate high light conditions in shallow waters), the modeled spacers fold to occlude the ribosomal binding site, while at lower temperatures (approximating deep water, low light conditions), a strongly conserved repeat facilitates the refolding of the transcript to fully expose the ribosomal binding site. Spacers lacking the 3-nt insert do not exhibit the conformational change when modeled under identical parameters but retain a fully functional structure across the same temperature gradient. From these observations, Peredo et al. (2012) concluded that *Najas* possesses a unique environmentally responsive mechanism for altering photosynthetic rates in different water depths, by regulating the degree of *psaB* translation, and thus PSI assembly (Fig. 4).

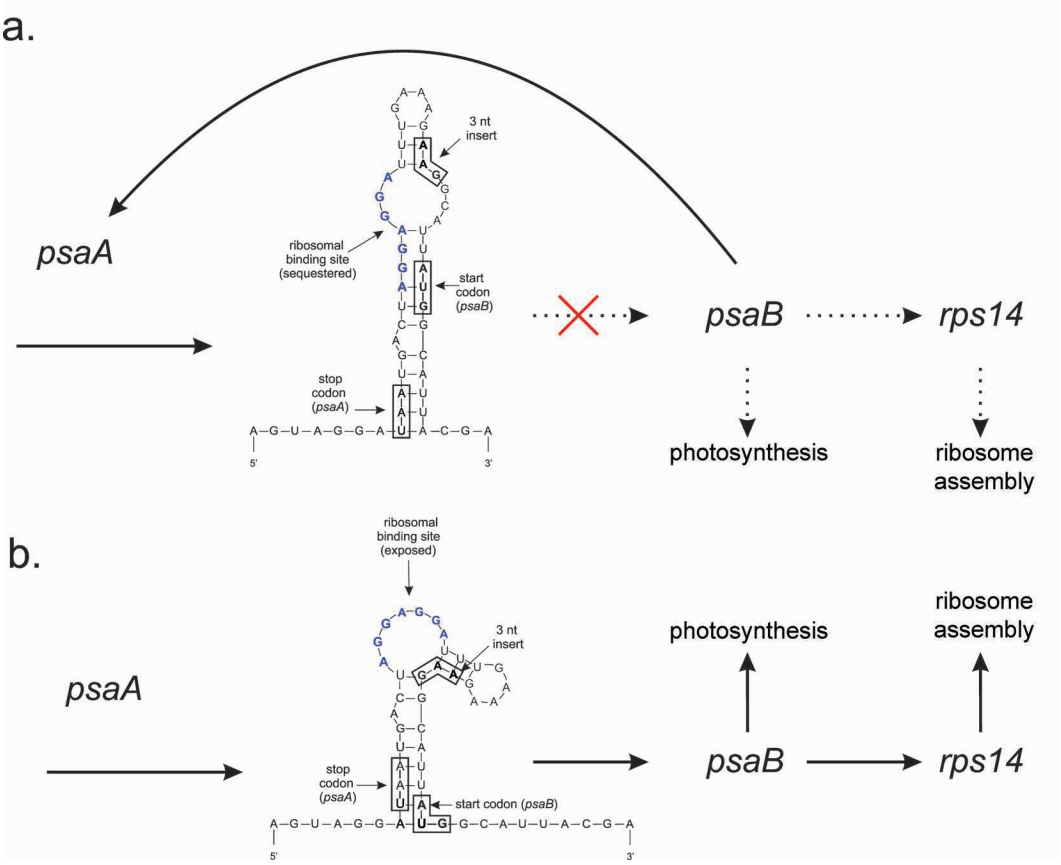


FIGURE 4. Proposed model of translational photosynthetic regulation in *Najas* (adapted from Peredo et al., 2013). In silico transcript modeling of the *psaA*/*psaB* spacer indicates that the 3-nt insert unique to *Najas* species induces a temperature-induced conformational change. **A.** In shallow waters, where temperatures are warm and light levels high, the spacer would fold to sequester the ribosomal binding site necessary for continued translation of *psaB*. The reduced translation of *psaB* would be balanced by downregulation of *psaA* because the stoichiometric ratio of both subunits must be equal for proper assembly of the critical photosystem I scaffold. By lowering production of photosystem I, the photosynthetic rate in shallow waters would be reduced. Because subsequent translation of the co-transcribed *rps14* is relative to that of *psaB*, the assembly of ribosomes predictably would decline as well. **B.** In deep waters, where temperatures and light levels are low, the *psaA*/*psaB* spacer conformation would fully expose the ribosomal binding site, thus allowing translation of *psaB* to proceed optimally following the expression of *psaA*. More equivalent expression of *psaA* and *psaB* subunits would yield greater quantities of photosytem I, leading to increased photosynthetic rates. Successful translation of *psaB* would enhance translation of *rps14*, resulting in increased ribosome assembly.

The cotranscription of one ribosomal protein (*rps14*) along with the core PSI components is of additional relevance. Due to the enormous metabolic investment in ribosome production (Warner, 1999), even small reductions in the rate of ribosome assembly or in the synthesis of individual ribosomal components should benefit the cell by minimizing energy wastage. Even subtle regulation of ribosome assembly can be critical in an evolutionary context (Li et al., 1996). Given that *rps14* is translated subsequent to *psaB*, its translation is dependent on that of *psaB*. Because *rps14* encodes the S14 protein incorporated into the 30s ribosomal subunit (Yamaguchi et al., 2000), any alteration of its expression predictably would result in associated perturbations of ribosome assembly. Consequently, temperature-mediated coregulation of *rps14* could represent a secondary adaptation associated with the modified *psaA*/*psaB* spacer of *Najas* by optimizing ribosome assembly in synchrony with the varying levels of PSI expression (Fig. 4).

Adaptive Loss of Gene Function in *Najas*: The Chloroplast *ndh* Complex

Since the chloroplast genome of higher plants was first sequestered from ancestral cyanobacterial symbionts, thousands of its genes were moved to the nuclear genome; however, the remaining cp genes represent a subset of those responsible for critical photosynthetic processes. By conducting a correlation analysis of chloroplast genomes, De Las Rivas et al. (2002) identified the most critical cp functions overall to be those involving bioenergetic protein complexes such as *psaA* and *psaB* [discussed above], *atpA* and *atpF*, and *rbcL*, as well as genes involved in transcription or translation (e.g., *rpl2*, *rps12*, *rpl16*, *rpoB*). In higher plants specifically, the most strongly correlated genes included *matK*, *petL*, *psbM*, and *ndhA-K*, the latter representing a suite of 11 NADH dehydrogenase genes (De Las Rivas et al., 2002).

Because of their ascribed chlororespiratory role, the *ndh* gene complex (NDH) is viewed as an important adaptation linked to terrestrial photosynthesis (Martín & Sabater, 2010). The NDH complex is believed to optimize rates of cyclic photophosphorylation and to alleviate photooxidative related stress induced by drought, low humidity, high light levels, and extreme high or low temperatures (Martín & Sabater, 2010; Yamori et al., 2011). Ecophysiological analyses have concluded that chlororespiration and the NDH complex are unnecessary under benign conditions but are indispensable when plants experience stress by preventing the accumulation of reactive oxygen species (Rumeau et al., 2007). The importance of *ndh* genes in averting photooxidative stress is illustrated by *Arbutus unedo* L., where the entire *ndhH-D* operon has been duplicated, possibly as a mechanism that enhances tolerance to excessively harsh conditions (Martínez-Alberola et al., 2013).

In particular, the *ndh* genes are thought to represent a critical adaptation to high and fluctuating light intensities that characterize terrestrial habitats (Martín & Sabater, 2010; Kouřil et al., 2014; Shikanai, 2014). Except for Charophytes and some Prasinophytes, *ndh* genes are absent in all algae, whose aquatic environment renders them less susceptible to photooxidative stress (Martín & Sabater, 2010). It is for this reason that the retention of *ndh* genes in Charophyte algae is viewed as an essential precursor to land plant evolution (Martín & Sabater, 2010). The vital function of *ndh* genes in angiosperms is evidenced by their absence in very few groups, such as nonphotosynthetic, parasitic genera like *Epifagus* and *Cuscuta* (Seliverstov et al., 2009). Some terrestrial angiosperms have gained pseudogenes (e.g., for *ndhF*) that arose by expansion of the inverted repeat; however, these represent duplications of functional *ndh* analogs (Seliverstov et al., 2009).

Although loss of gene function once was regarded as maladaptive, evidence now suggests that overrepresented loss of function in enzymes and regulatory proteins often reflects altered regulatory and metabolic pathways, which facilitate adaptation to novel environmental conditions (Hottes et al., 2013). Because their function is tied so closely to terrestrial conditions, the NDH complex seems to be a likely candidate for loss of function in aquatic plants as a consequence of relaxed selective pressures to retain it.

Iles et al. (2013) hypothesized multiple, convergent *ndh* gene losses in alismatid angiosperms after detecting loss or pseudogenization of *ndhB* and *ndhF* in five of 21 genera surveyed (*Amphibolis*, *Najas*, *Posidonia*, *Thalassia*, *Vallisneria*). Although this group comprises aquatic and wetland species entirely (Les & Tippery, 2013), functional *ndhB* and *ndhF* genes have been retained by the majority (Iles et al., 2013). More specific details of *ndh* gene distribution in alismatid monocots became available when complete chloroplast genome sequences were published for two genera. The cp genome of *Elodea* (Huotari & Korpelainen, 2012) contains functional sequences for all 11 plastid *ndh* genes, including the two reported by Iles et al. (2013). However, in *Najas* (Peredo et al., 2013), six *ndh* genes are absent and the remaining five (including *ndhB* and *ndhF*) have been converted to pseudogenes (Fig. 5).

Convergent losses of *ndh* genes in different alismatid clades (Iles et al., 2013) attest to the adaptive nature of this feature. However, a broader evaluation of submersed aquatic plant genera for which data are available indicates that at least some *ndh* genes remain intact in most of them (Table 3), and that the only known losses are restricted phylogenetically to the alismatids. If the loss of *ndh* gene function is an adaptation arising from reduced selection in aquatic environments, then it may seem unusual that many submersed aquatic plants have retained the complete functional set of genes, nevertheless. Several explanations of this perceived anomaly can be offered.

It always is possible that many of those aquatic plants retaining functional cp *ndh* genes are no longer under selection to do so, but simply have not yet encountered an opportunity to lose them. Such an explanation is suggested by the erratic loss of adaptive secondary metabolites in the submersed foliage of different heterophyllous *Potamogeton* species (Les & Sheridan, 1990). However, this explanation does not adequately explain the retention of *ndh* genes in genera like *Ceratophyllum*, which is estimated to have

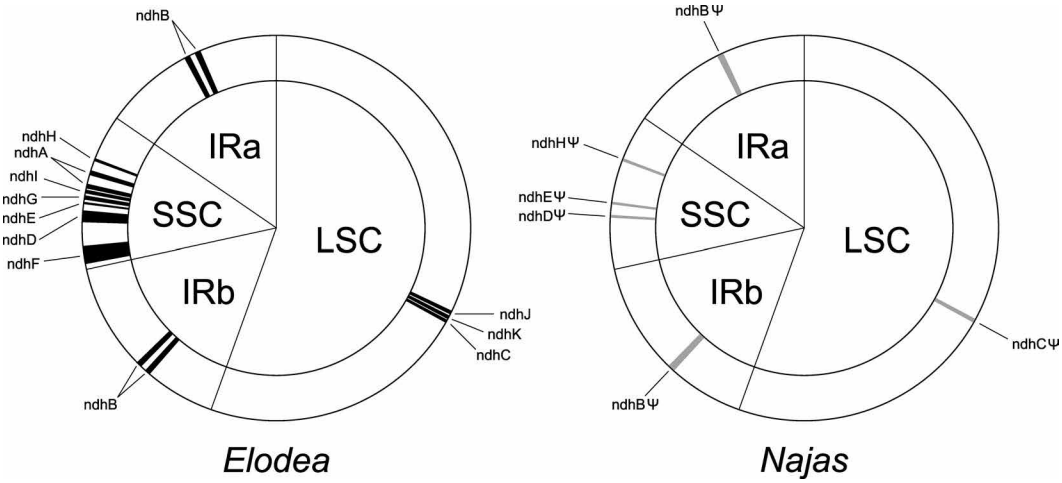


FIGURE 5. Comparison of *ndh* genes based on complete chloroplast genome sequences of *Elodea* (Huotari & Korpelainen, 2012) and *Najas* (Peredo et al., 2013). *Elodea* (left) retains a fully functional complement of *ndh* genes (black); however, *Najas* has lost six *ndh* genes (*ndhA*, *ndhF*, *ndhG*, *ndhI*, *ndhJ*, *ndhK*), while five others (*ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhH*) have been converted to pseudogenes (gray). Ψ = pseudogenes; indicated schematically: LSC = large single-copy region; SSC = small single-copy region; IR = inverted repeats.

diverged some 140 million years ago (Moore et al., 2007), more than sufficient time for gene loss or at least pseudogenization to have occurred. The pattern of loss within alismatids also indicates that *ndh* stop codons have been generated principally by frameshifts resulting from nontriplet indels (Iles et al., 2013), a mechanism that would produce pseudogenes quite readily.

The mosaic pattern of *ndh* gene loss in submersed aquatic plants possibly reflects mechanistic factors. As one example, the greater conservation of the *ndhF* promoter in dicots than in monocots (Seliverstov et al., 2009) might result in the latter group being more susceptible to *ndh* gene loss overall. Once a promoter loses function, the dissolution of the associated gene(s) is likely to occur rapidly. Interestingly, all of the known cp *ndh* gene losses in submersed aquatic plants occur in the monocots, with apparent gene retention in all submersed dicots surveyed (Table 3). Despite there being 11 *ndh* genes, their arrangement in only four transcription units: *ndhF*, *ndhB*, *ndhH*/*A*/*I*/*G*/*E*/*D*, and *ndhC*/*K*/*J* (Suorsa et al., 2009), could simplify dismantling of the entire NDH complex by requiring fewer promoter mutations. Furthermore, the inactivation of only a single *ndh* gene is known to suppress the expression of the whole complex. A documented example involves the mechanism of RNA editing, which is known to occur in all 11 *ndh* genes (Brennicke et al., 2014). In this case, *Arabidopsis* mutants having defective RNA editing of only the *ndhD* initiation codon exhibit extensively reduced accumulation of the entire NDH complex (Kotera et al., 2005). Conceivably, the disruption of a single *ndh* gene, or its promoter, by any means might be expected to have similar consequences.

It is entirely plausible that the modified *psaA*/*psaB* spacer of *Najas* described above represents another mechanistic factor that somehow is associated with the loss of *ndh* genes, at least in this genus. Recent studies indicate that PSI and *ndh* genes together form a "supercomplex," which is essential for maintaining stability of the NDH complex, especially under conditions of high light stress (Peng & Shikanai, 2011). Because of this close association between PSI and the NDH complex, it is feasible that any adaptation capable of regulating PSI to reduce photosynthetic rates in shallow, high light water environments could obviate the need to retain *ndh* genes due to relaxed selection pressure.

Conceivably, this association could explain why *Najas*, which possesses the modified *psaA*/*psaB* spacer, has lost its *ndh* gene function, while the closely related *Elodea*, which possesses an unmodified spacer, also retains its full complement of *ndh* genes (Fig. 5). Otherwise, these genera exhibit no obvious ecological differences. Both genera frequently cohabit similar aquatic environments; with *Elodea* actually occupying a slightly broader range of depths than *Najas* (Nichols, 1999). The life-history of the two genera does differ, in that *Najas* is annual (a habit otherwise rare among submersed aquatic plants) and *Elodea* perennial.

TABLE 3. Distribution of chloroplast *ndh* genes in 22 submersed aquatic plant genera compiled from data in GenBank.

Taxon	*ndhA*	*ndhB*	*ndhC*	*ndhD*	*ndhE*	*ndhF*	*ndhG*	*ndhH*	*ndhI*	*ndhJ*	*ndhK*
Charophyta											
Characeae											
Chara	+	+	+	+	+	+	+	+	+	+	+
Lycopodiophyta											
Isoetaceae											
Isoëtes	+	+	+	+	+	+	+	+	+	+	+
Miscellaneous dicots											
Cabombaceae											
Cabomba	n/a	+	n/a	n/a	n/a	+	n/a	n/a	n/a	n/a	n/a
Ceratophylaceae											
Ceratophyllum	+	+	+	+	+	+	+	+	+	+	+
Alismatid monocots											
Aponogetonaceae											
A*ponogeton*	n/a	n/a	n/a	n/a	n/a	+	n/a	n/a	n/a	n/a	n/a
Cymodoceaceae											
Amphibolis	n/a	ψ	n/a	n/a	n/a	ψ	n/a	n/a	n/a	n/a	n/a
Cymodoceaceae											
Halodule	n/a	+	n/a	n/a	n/a	+	n/a	n/a	n/a	n/a	n/a
Hydrocharitaceae											
Thalassia	n/a	ψ	n/a	n/a	n/a	–	n/a	n/a	n/a	n/a	n/a
Vallisneria	n/a	n/a	n/a	n/a	n/a	ψ	n/a	n/a	n/a	n/a	n/a
Elodea	+	+	+	+	+	+	+	+	+	+	+
Najas	–	ψ	ψ	ψ	ψ	–	–	ψ	–	–	–
Posidoniaceae											
Posidonia	n/a	ψ	n/a	n/a	n/a	–	n/a	n/a	n/a	n/a	n/a
Ruppiaceae											
Ruppia	n/a	+	n/a	n/a	n/a	+	n/a	n/a	n/a	n/a	n/a
Zannichelliaceae											
Zannichellia	n/a	+	n/a	n/a	n/a	+	n/a	n/a	n/a	n/a	n/a
Zosteraceae											
Zostera	n/a	+	n/a	n/a	n/a	+	n/a	n/a	n/a	n/a	n/a
Eudicots											
Elatinaceae											
Elatine	n/a	n/a	n/a	n/a	n/a	+	n/a	n/a	n/a	n/a	n/a
Haloragaceae											
Myriophyllum	n/a	n/a	n/a	n/a	n/a	+	n/a	+	n/a	n/a	n/a
Lentibulariaceae											
Utricularia	+	+	+	+	+	+	+	+	+	+	+
Plantaginaceae											
Hippuris	n/a	n/a	n/a	n/a	n/a	+	n/a	n/a	n/a	n/a	n/a
Limnophila	n/a	n/a	n/a	n/a	n/a	+	n/a	+	n/a	n/a	n/a
Podostomaceae											
Podostemum	+	+	+	+	+	+	+	+	+	+	+
Pontederiaceae											
Heteranthera	n/a	n/a	n/a	n/a	n/a	+	n/a	n/a	n/a	n/a	n/a

Although these taxa inhabit a comparable environment, losses of a functional NDH complex have been documented only in alismatid monocots. Currently, only sporadic information is available for many taxa and comprehensive evaluations are possible only for those genera (in bold) where full chloroplast genomes have been sequenced.
+ = gene present; – = gene absent; ψ = pseudogene only; n/a = not surveyed.

The irregular distribution of nonfunctional *ndh* genes in submersed aquatic plants indicates that some species likely remain under selection to retain *ndh* genes because of specific ecological constraints. At least some genera that retain the NDH complex often grow near the water surface where light intensities consistently are comparable to terrestrial conditions. The rootless condition of *Ceratophyllum* and *Utricularia* results in a "suspended" habit where the plants float just beneath the water surface (Hutchinson, 1975). The riverine habitat of *Podostemum* subjects the plants to full immersion during part of the season, but fully exposes them to sunlight as water levels drop seasonally and flowering commences (Philbrick & Novelo, 2004). Understandably, the retention of *ndh* genes by these species does not seem unusual in light of their ecology. In contrast, those species where *ndh* genes have been lost tend to occur nearly exclusively beneath the water surface. These include the freshwater *Najas* and *Vallisneria*, as well as *Amphibolis*, *Posidonia*, and *Thalassia*, which occupy oceanic habitats as submersed seagrasses.

Positive Selection in Aquatic Plants

With few exceptions, widespread adaptive amino acid substitution has been documented rarely in plant taxa (Slotte et al., 2010). It can be difficult to demonstrate positive selection in specific taxa due to the overall rarity of positively selected sites in the genome and because of the fairly conservative nature of the tests available. Genes overwhelmingly have experienced purifying selection, which removes deleterious nonsynonymous mutations and results in an ω ratio < 1 (Luo & Hughes, 2012). Thus, even when thousands of genes are screened, the possibility of positive selection often is indicated for only a few percent (Sun et al., 2013).

Indications of selection also require appropriate models that take into account such factors as base/codon frequency bias (Luo & Hughes, 2012). To be most informative, the models also must be able to identify patterns of selection at individual sites or in particular taxa. Earlier branch and site models (Yang, 1998; Yang et al., 2000; Yang & Nielsen, 2002) evaluated positive selection from significant departures of ω but were based on average d_N/d_S ratios over all branches or all sites respectively. The branch-site model (Yang et al., 2005; Zhang et al., 2005), which we used here, provides much better resolution and is able to detect positive selection in cases in which only a few sites on specified lineages are affected. Although much improved over earlier models, the branch-site model test remains conservative in analyses in which saturation is a consideration (Gharib & Robinson-Rechavi, 2013). Saturation becomes increasingly problematic as comparisons are expanded to encompass broad phylogenetic distances. As a consequence, it is relatively difficult to identify sites under positive selection, in a specific taxon and across a broad phylogenetic background. Significant results in such cases should be viewed as providing particularly strong evidence of adaptive substitutions.

Given the adaptive heritage of terrestrial plant metabolic enzymes, however, it is reasonable to expect that they might not be able to function effectively across the broad temperature gradient tolerated by submersed hydrophytes (Fig. 1). Thus, it may be expected that at least some of the metabolic enzymes of aquatic plants would have required significant adaptive modifications necessary to maintain their physiological function under different environmental extremes. This prediction is borne out to some degree by evolutionary studies of aquatic mammals, where a search for candidate genes relating to aquatic adaptations in dolphins found that 3.1% of the nearly 12 000 genes evaluated had undergone positive selection (Sun et al., 2013). That group of positively selected genes included many loci that were identified as relating to potential aquatic adaptations in those cetaceans.

Although a comparable genome-wide analysis has not yet been carried out for submersed aquatic plants, several studies have suggested the possibility that positive selection has acted on at least some genes. The cp *rbcL* gene is believed to have evolved under positive selection in some species of the aquatic angiosperm genus *Potamogeton* (Iida et al., 2009). Signals of positive selection also have been indicated for the cp *rbcL*+intergenic spacer region and for the nuclear *EF-1*α gene of the macroalga *Chara braunii* C. C. Gmel. (Kato et al., 2011). Interestingly, the key elongation factor involved in protein translation is represented by the gene *EF-1*α in Charophytes and land plants, but by an analog known as EFL (elongation factor-like) in many other algae (Cocquyt et al., 2009). Thus, the potential modification of elongation factors as aquatic adaptations deserves further inquiry, especially for those species capable of maintaining translation across a broad, depth-related temperature gradient.

Evidence of positive selection in *Najas* was sought first by evaluating the recently sequenced cp genome, which is a promising target of selection in aquatic plants due to its primary dedication to photosynthetic processes. A comparison of cp genomes from *Najas* and the related *Elodea* (both Hydrocharitaceae) disclosed several areas where the former had diverged more extensively from terrestrial angiosperms than had the latter (Peredo et al., 2013). Although one

anomalous pattern clearly reflected the absence and pseudogenization of multiple *ndh* genes in *Najas*, another divergent area was indicated for the operon that codes genes for B and C subunits of the PEP. We focused on this region for several other reasons. Steiner et al. (2011) claimed that PEP was the "predominant target for environmental regulation" in the chloroplast, where it is involved in light-induced control of transcription (Steiner et al. 2011: 1044). Furthermore, because the *psaA*/*psaB*/*rps14* operon (described above) contains three of the cp genes that rely exclusively on PEP for transcription (Lopez Peredo et al., 2012), it was reasonable to expect that modifications to that RNA polymerase might be necessary to facilitate transcription across the entire temperature and light regime associated with the regulation of PSI translation by the modified *psaA*/*psaB* spacer of *Najas*.

The complete PEP enzyme comprises four cp-encoded subunits: *rpoA*, *rpoB*, $rpoC_1$, and $rpoC_2$ (Serino & Maliga, 1998; Steiner et al., 2011), which may assemble with various nuclear-encoded, promoter-specific sigma factors (Hanaoka et al., 2005). The *rpoA* subunit is involved with complex stabilization, the *rpoB* subunit with RNA synthesis, and the $rpoC_2$ subunit with DNA binding; the function of the $rpoC_1$ remains uncertain (Steiner et al., 2011). Understandably, *rpoB* has been identified as carrying out one of the most critical functions of the cp genome (De Las Rivas et al., 2002).

Evaluation of d_N/d_S ratios in all four PEP subunits of *Najas* disclosed evidence of positive selection in two of them: *rpoB* and $rpoC_1$ (Table 2). Although evidence of positive selection can be difficult to obtain, especially in comparisons over wide phylogenetic distances, the significant ω values obtained in these comparisons were particularly compelling for several reasons. Given the lack of selection observed in the closely related genus *Elodea*, divergence of *rpoB* and $rpoC_1$ in *Najas* does not simply reflect phylogenetic distance, nor does it broadly associate selection at these loci with other water plants. In other words, the changes in PEP that occur in *Najas* appear to be specific to that genus. Furthermore, not only do the same selected sites occur in all *Najas* species examined across the phylogenetic breadth of the genus, the few point mutations observed at these sites were synonymous (i.e., they retained the same divergent amino acid substitutions).

The significantly elevated ω values in the PEP operon that were observed only in *Najas* but not in the closely related *Elodea*, comprised a result predicted by a BLAST (Basic Local Alignment Search Tool, NCBI; http://blast.ncbi.nlm.nih.gov/Blast.cgi) comparison of *Elodea* and *Najas* cp genome coding sequences relative to other angiosperms (Peredo et al., 2013). Also, modifications to PEP, enabling effective function under various light and temperature regimes, predictably would be necessary to accommodate transcription of the modified *psaA*/*psaB*/*rps14* operon in *Najas*, which relies exclusively on this polymerase.

There is little doubt that PEP has in some way been modified functionally in *Najas* relative to *Elodea* and other angiosperms. In particular, the *rpoB* subunit (critically responsible for RNA synthesis) is extremely divergent in *Najas* when compared with a phylogenetically diverse selection of angiosperms. The large number (10) of significant nonsynonymous substitutions in this gene, coupled with a concentration of five substitutions within a nearly contiguous section of the gene, and three of these changing amino acid charges (Fig. 3), surely suggests some level of metabolic modification. Even more radical substitutions occur within the $rpoC_1$ subunit (Fig. 3), for which a specific function unfortunately has not yet been elucidated. In considering this extent of nonsynonymous and charge-altering amino acid substitutions in two of its four subunits, it is difficult to envision that PEP would not function differently in *Najas*. Although the functional ramifications of PEP modification in *Najas* will require experimental elucidation, it is intriguing to contemplate whether the altered amino acid composition of the enzyme might somehow facilitate transcription across the same temperature extremes to which the translational regulatory function of the modified *psaA*/*psaB* spacer is indicated.

Despite *Najas* and *Elodea* occupying very similar ecological habitats, the lack of positive selection in any PEP subunits in the latter genus indicates that comparable enzymatic modification has been unnecessary. Because *Elodea* also retains the NDH complex and possesses an unmodified *psaA*/*psaB* spacer, we are led to hypothesize that that spacer modification, NDH loss, and modified PEP structure may represent interrelated photosynthetic adaptations in *Najas*.

Conclusions

Although the arguments presented in this study have enabled us to provide cogent hypotheses of photosynthetic adaptation in the highly specialized aquatic genus *Najas*, further testing of the proposed concepts is necessary. In particular, it would be highly informative to evaluate the PEP variant of *Najas* in vitro across a broad environmental light and temperature gradient to determine to what extent the metabolic properties of the modified enzyme have been altered.

The PEP variant of *Elodea* would provide a suitable comparison in this case because of its close phylogenetic relationship and similar amino acid composition to terrestrial angiosperms.

Further efforts should be made to investigate the influence of selection on other plastid as well as nuclear genes of aquatic angiosperms. In particular, there has been increased appreciation that the inappropriate inclusion of gene sequences from unusually evolving loci can result in misleading phylogenetic inferences (Novis et al., 2013). Therefore, a more comprehensive evaluation of selection in aquatic plant genomes should provide better insight into determining which loci are most appropriate (or inappropriate) to use in the phylogenetic analysis of water plants.

Acknowledgments

Portions of this work were supported by grants from the National Science Foundation (NSF DEB–0841658); Fulbright Foundation and Spanish Ministerio de Educación, Cultura y Deporte; Irish Research Council for Science, Engineering, and Technology; and the University of Connecticut College of Liberal Arts and Sciences.

Literature Cited

[APG III] The Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society 161: 105–121. https://doi.org/10.1111/j.1095-8339.2009.00996.x

Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell & D. L. Wheeler. 2005. GenBank. Nucleic Acids Research, Supplement I 33: D34–D38. https://doi.org/10.1093/nar/gki063

Bock, R. & J. N. Timmis. 2008. Reconstructing evolution: gene transfer from plastids to the nucleus. BioEssays 30: 556–566. https://doi.org/10.1002/bies.20761

Brennicke, A., E. Picardi, C. Quagliariello, & T. M. R. Regina. 2014. REDIdb, the RNA editing database; http://biologia.unical.it/py_script/overview.html [accessed 4 February 2014].

Choquet, Y. & F.-A. Wollman. 2002. Translational regulations as specific traits of chloroplast gene expression. FEBS Letters 529: 39–42. https://doi.org/10.1016/s0014-5793(02)03260-x

Cocquyt, E., H. Verbruggen, F. Leliaert, F. W. Zechman, K. Sabbe, & O. De Clerck. 2009. Gain and loss of elongation factor genes in green algae. BMC Evolutionary Biology 9: 39. https://doi.org/10.1186/1471-2148-9-39

Cook, C. D. K. 1996. Aquatic Plant Book, 2nd rev. ed. SPB Academic Publishing, Amsterdam, The Netherlands.

———. 1999. The number and kinds of embryo-bearing plants which have become aquatic: a survey. Perspectives in Plant Ecology, Evolution and Systematics 2: 79–102. https://doi.org/10.1078/1433-8319-00066

De Las Rivas, J., J. J. Lozano, & A. R. Ortiz. 2002. Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns. Genome Research 12: 567–583. https://doi.org/10.1101/gr.209402

Delaux, P.-M., A. K. Nanda, C. Mathé, N. Sejalon-Delmas, & C. Dunand. 2012. Molecular and biochemical aspects of plant terrestrialization. Perspectives in Plant Ecology, Evolution and Systematics 14: 49–59. https://doi.org/10.1016/j.ppees.2011.09.001

DePamphilis, C. W. & J. D. Palmer. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348: 337–339. https://doi.org/10.1038/348337a0

Drummond, A. J., B. Ashton, S. Buxton, M. Cheung, A. Cooper, J. Heled, M. Kearse, R. Moir, S. Stones-Havas & S. Sturrock. 2010. Geneious version 6.1.6; http://geneious.com.

Evert, R. F. & S. E. Eichhorn. 2012. Raven Biology of Plants, 8th ed. W. H. Freeman and Company, Publishers, New York, NY.

Fassett, N. C. 1957. A Manual of Aquatic Plants. With revision appendix by E. C. Ogden. The University of Wisconsin Press, Madison, WI.

Filatov, V., J. Dowdle, N. Smirnoff, B. Ford-Lloyd, H. J. Newbury, & M. R. MacNair. 2006. Comparison of gene expression in segregating families identifies genes and genomic regions involved in a novel adaptation, zinc hyperaccumulation. Molecular Ecology 15: 3045–3059. https://doi.org/10.1111/j.1365-294x.2006.02981.x

Gharib, W. H. & M. Robinson-Rechavi. 2013. The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. Molecular Biology and Evolution 30: 1675–1686. https://doi.org/10.1093/molbev/mst062

Hanaoka, M., K. Kanamaru, M. Fujiwara, H. Takahashi & K. Tanaka. 2005. Glutamyl-tRNA mediates a switch in RNA polymerase use during chloroplast biogenesis. EMBO Reports 6: 545–550. https://doi.org/10.1038/sj.embor.7400411

Hottes, A. K., P. L. Freddolino, A. Khare, Z. N. Donnell, J. C. Liu, & S. Tavazoie. 2013. Bacterial adaptation through loss of function. PLoS Genetics 9, e1003617. https://doi.org/10.1371/journal.pgen.1003617

Huotari, T. & H. Korpelainen. 2012. Complete chloroplast genome sequence of Elodea canadensis and comparative analyses with other monocot plastid genomes. Gene 508: 96–105. https://doi.org/10.1016/j.gene.2012.07.020

Hutchinson, G. E. 1975. A Treatise on Limnology. Volume 3: Limnological Botany. John Wiley & Sons, New York, NY.

Iida, S., A. Miyagi, S. Aoki, M. Ito, Y. Kadono, & K. Kosuge. 2009. Molecular adaptation of rbcL in the heterophyllous aquatic plant Potamogeton. PLoS ONE 4: e4633. https://doi.org/10.1371/journal.pone.0004633

Iles, W. J. D., S. Y. Smith & S. W. Graham. 2013. A well-supported phylogenetic framework for the monocot order Alismatales reveals multiple losses of the plastid NADH dehydrogenase complex and a strong long-branch effect. Pp. 1–28 *in:* Wilkin, P. & S. J. Mayo, Eds. Early Events in Monocot Evolution. Cambridge University Press, New York, NY. https://doi.org/10.1017/cbo9781139002950.002

Kato, S., K. Misawa, F. Takahashi, H. Sakayama, S. Sano, K. Kosuge, F. Kasai, M. M. Watanabe, J. Tanaka & H. Nozaki. 2011. Aquatic plant speciation affected by diversifying selection of organelle DNA regions. Journal of Phycology 47: 999–1008. https://doi.org/10.1111/j.1529-8817.2011.01037.x

Kotera, E., M. Tasaka & T. Shikanai. 2005. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature 433: 326–330. https://doi.org/10.1038/nature03229

Kouřil, R., O. Strouhal, L. Nosek, R. Lenobel, I. Chamrád, E. J. Boekema, M. Šebela & P. Ilík. 2014. Structural characterization of a plant photosystem I and NAD(P)H dehydrogenase supercomplex. The Plant Journal 77: 568–576. https://doi.org/10.1111/tpj.12402

Kuroda H. & P. Maliga. 2003. The plastid clpP1 protease gene is essential for plant development. Nature 425: 86–89. https://doi.org/10.1038/nature01909

Les, D. H. 1988. Breeding systems, population structure, and evolution in hydrophilous angiosperms. Annals of the Missouri Botanical Garden 75: 819–835. https://doi.org/10.2307/2399370

———& D. J. Sheridan. 1990. Biochemical heterophylly and flavonoid evolution in North American Potamogeton (Potamogetonaceae). American Journal of Botany 77: 453–465. https://doi.org/10.2307/2444379

———, R. S. Capers & N. P. Tippery. 2006. Introduction of Glossostigma (Phrymaceae) to North America: a taxonomic and ecological overview. American Journal of Botany 93: 927–939. https://doi.org/10.3732/ajb.93.6.927

———, S. P. Sheldon & N. P. Tippery. 2010. Hybridization in hydrophiles: natural interspecific hybrids in Najas L. (Hydrocharitaceae). Systematic Botany 35: 736–744. https://doi.org/10.1600/036364410x539826

Li, B., J. Vilardell & J. R. Warner. 1996. An RNA structure involved in feedback regulation of splicing and of translation is critical for biological fitness. Proceedings of the National Academy of Sciences of the U.S.A. 93: 1596–1600. https://doi.org/10.1073/pnas.93.4.1596

Luo, H. & A. L. Hughes. 2012. dN/dS does not show positive selection drives separation of polar-tropical SAR11 populations. Molecular Systems Biology 8: 625. https://doi.org/10.1038/msb.2012.58

Marín-Navarro, J., A. L. Manuell, J. Wu & S. P. Mayfield. 2007. Chloroplast translation regulation. Photosynthesis Research 94: 359–374. https://doi.org/10.1007/s11120-007-9183-z

Martín, M. & B. Sabater. 2010. Plastid ndh genes in plant evolution. Plant Physiology and Biochemistry 48: 636–645. https://doi.org/10.1016/j.plaphy.2010.04.009

Martínez-Alberola, F., E. M. del Campo, D. Lázaro-Gimeno, S. Mezquita-Claramonte, A. Molins, I. Mateu-Andrés, J. Pedrola-Monfort, L. M. Casano & E. Barreno. 2013. Balanced gene losses, duplications and intensive rearrangements led to an unusual regularly sized genome in Arbutus unedo chloroplasts. PLoS ONE 8: e79685. https://doi.org/10.1371/journal.pone.0079685

McFadden, G. I. 2001. Chloroplast origin and integration. Plant Physiology 125: 50–53. https://doi.org/10.1104/pp.125.1.50

Meng, B. Y., M. Tanaka, T. Wakasugi, M. Ohme, K. Shinozaki & M. Sugiura. 1988. Cotranscription of the genes encoding two P700 chlorophylla apoproteins with the gene for ribosomal protein CS14: determination of the transcriptional initiation site by in vitro capping. Current Genetics 14: 395–400 https://doi.org/10.1007/bf00419998

Midorikawa, T., K. Matsumoto, R. Narikawa & M. Ikeuchi. 2009. An Rrf2-type transcriptional regulator is required for expression of psaAB genes in the Cyanobacterium *Synechocystis* sp. PCC 68031. Plant Physiology 151: 882–892. https://doi.org/10.1104/pp.109.141390

Moore, M. J., C. D. Bell, P. S. Soltis & D. E. Soltis. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proceedings of the National Academy of Sciences of the U.S.A. 104: 19363–19368. https://doi.org/10.1073/pnas.0708072104

Nichols, S. A. 1999. Distribution and habitat descriptions of Wisconsin lake plants. Bulletin 96. Wisconsin Geological & Natural History Survey, Madison, WI.

Novis, P. M., R. Smissen, T. R. Buckley, K. Gopalakrishnan & G. Visnovsky. 2013. Inclusion of chloroplast genes that have undergone expansion misleads phylogenetic reconstruction in the *Chlorophyta*. American Journal of Botany 100: 2194–2209. https://doi.org/10.3732/ajb.1200584

Ogden, E. C. 1974. Anatomical patterns of some aquatic vascular plants of New York. Bulletin No. 424. New York State Museum and Science Service, Albany, NY.

Pedersen, O., T. D. Colmer & K. Sand-Jensen. 2013. Underwater photosynthesis of submerged plants—

recent advances and methods. Frontiers in Plant Science 4: 140. https://doi.org/10.3389/fpls.2013.00140

Peng, L. & T. Shikanai. 2011. Supercomplex formation with photosystem I is required for the stabilization of the chloroplast NADH dehydrogenase-like complex in Arabidopsis. Plant Physiology 155: 1629–1639. https://doi.org/10.1104/pp.110.171264

Peredo, E., D. H. Les, U. M. King & L. K. Benoit. 2012. Extreme conservation of the psaA/psaB intercistronic spacer reveals a translational motif coincident with the evolution of land plants. Journal of Molecular Evolution 75: 184–197. https://doi.org/10.1007/s00239-012-9526-z

———, U. M. King & D. H. Les. 2013. The chloroplast genome of Najas flexilis: adaptation to submersed environments is accompanied by the complete loss of the NDH complex in an aquatic angiosperm. PLoS ONE 8: e68591.

Philbrick, C. T. & A. Novelo R. 2004. Monograph of *Podostemum* (Podostemaceae). Systematic Botany Monographs 70: 1–106. https://doi.org/10.2307/25027922

Rumeau, D., G. Peltier & L. Cournac. 2007. Chlororespiration and cyclic electron flow around PSI during photosynthesis and plant stress response. Plant, Cell & Environment 30: 1041–1051. https://doi.org/10.1111/j.1365-3040.2007.01675.x

Schwabe, T. M. & J. Kruip. 2000. Biogenesis and assembly of photosystem I. Indian Journal of Biochemistry and Biophysics 37: 351–359.

Sculthorpe, C. D. 1967. The Biology of Aquatic Vascular Plants. Edward Arnold (Publishers) Ltd., London, U.K.

Seliverstov, A. V., E. A. Lysenko & V. A. Lyubetsky. 2009. Rapid evolution of promoters for the plastome gene ndhF in flowering plants. Russian Journal of Plant Physiology 56: 838–845. https://doi.org/10.1134/s1021443709060144

Semenov, A. Y., I. V. Shelaev, F. E. Gostev, M. D. Mamedov, V. A. Shuvalov, O. M. Sarkisov & V. A. Nadtochenko. 2012. Primary steps of electron and energy transfer in photosystem I: effect of excitation pulse wavelength. Biochemistry (Moscow) 77: 1011–1020. https://doi.org/10.1134/s0006297912090088

Serino, G. & P. Maliga. 1998. RNA polymerase subunits encoded by the plastid *rpo* genes are not shared with the nucleus-encoded plastid enzyme. Plant Physiology 117: 1165–1170. https://doi.org/10.1104/pp.117.4.1165

Shikanai, T. 2014. Central role of cyclic electron transport around photosystem I in the regulation of photosynthesis. Current Opinion in Biotechnology 26: 25–30. https://doi.org/10.1016/j.copbio.2013.08.012

Slotte, T., J. P. Foxe, K. M. Hazzouri & S. I. Wright. 2010. Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. Molecular Biology and Evolution 27: 1813–1821. https://doi.org/10.1093/molbev/msq062

Steiner, S., Y. Schröter, J. Pfalz & T. Pfannschmidt. 2011. Identification of essential subunits in the plastid-encoded RNA polymerase complex reveals building blocks for proper plastid development. Plant Physiology 157: 1043–1055. https://doi.org/10.1104/pp.111.184515

Sun, Y.-B., W.-P. Zhou, H.-Q. Liu, D. M. Irwin, Y.-Y. Shen & Y. P. Zhang. 2013. Genome-wide scans for candidate genes involved in the aquatic adaptation of dolphins. Genome Biology and Evolution 5: 130–139. https://doi.org/10.1093/gbe/evs123

Suorsa, M., S. Sirpio & E.-M. Aro. 2009. Towards characterization of the chloroplast NAD(P)H dehydrogenase complex. Molecular Plant 2: 1127–1140. https://doi.org/10.1093/mp/ssp052

Swain, T. 1970. Biochemical Evolution and the Origin of Life. North-Holland Publishing Co., Amsterdam, The Netherlands.

———. 1975. Evolution of flavonoid compounds. Pp. 1096–1129 *in:* J. B. Harborne, T. J. Mabry, & H. Mabry, Eds. The Flavonoids. Academic Press, New York, NY. https://doi.org/10.1007/978-1-4899-2909-9_20

Vercauteren, N, H. Huwald, E. Bou-Zeid, J. S. Selker, U. Lemmin, M. B. Parlange & I. Lunati. 2011. Evolution of superficial lake water temperature profile under diurnal radiative forcing. Water Resources Research 47: W09522. https://doi.org/10.1029/2011wr010529

Warner, J. R. 1999. The economics of ribosome biosynthesis in yeast. Trends in Biochemical Sciences 24: 437–440. https://doi.org/10.1016/s0968-0004(99)01460-7

West-Eberhard, M. J. 2008. Toward a modern revival of Darwin's theory of evolutionary novelty. Philosophy of Science 75: 899–908. https://doi.org/10.1086/594533

Wostrikoff, K., J. Girard-Bascou, F.-A. Wollman & Y. Choquet. 2004. Biogenesis of PSI involves a cascade of translational autoregulation in the chloroplast of *Chlamydomonas*. Embo Journal 23: 2696–2705. https://doi.org/10.1038/sj.emboj.7600266

Yamaguchi, K., K. von Knoblauch & A. R. Subramanian. 2000. The plastid ribosomal proteins: identification of all the proteins in the 30 S subunit of an organelle ribosome (chloroplast). The Journal of Biological Chemistry 275: 28455–28465. https://doi.org/10.1074/jbc.m004350200

Yamori, W., N. Sakata, Y. Suzuki, T. Shikanai & A. Makino. 2011. Cyclic electron flow around photosystem I via chloroplast NAD(P)H dehydrogenase (NDH) complex performs a significant physiological role during photosynthesis and plant growth at low temperature in rice. The Plant Journal 68: 966–976. https://doi.org/10.1111/j.1365-313x.2011.04747.x

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular Biology and Evolution 15: 568–573. https://doi.org/10.1093/oxfordjournals.molbev.a025957

———. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24: 1586–1591. https://doi.org/10.1093/molbev/msm088

———& R. Nielsen. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Molecular Biology and Evolution 19: 908–917. https://doi.org/10.1093/oxfordjournals.molbev.a004148

———, ———, N. Goldman & A. M. Pedersen. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155: 431–449.

———, W. S. Wong & R. Nielsen. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Molecular Biology and Evolution 22: 1107–1118. https://doi.org/10.1093/molbev/msi097

———, ———& Z. Yang. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Molecular Biology and Evolution 22: 2472–2479. https://doi.org/10.1093/molbev/msi237

Printed from NYBG Press Website. Published by NYBG Press, 2900 Southern Boulevard, Bronx, NY 10458-5126

Published Online: 23 June 2017

NYBG PRESS

Chapter 3

Plastid genome sequencing of nine *Najas* taxa, and a partial plastome sequence for *Hydrilla verticillata* provides evidence of atypical plastome evolution in these aquatic plants

Abstract

The majority of angiosperm plastomes are conserved in size, gene content, order and structure; however, a small number of lineages are characterized by unusual plastome evolution. In this study, nine newly sequenced *Najas* plastomes are provided, along with a partial plastome assembly for *Hydrilla verticillata*.

Results show that the plastomes of *Najas* and *Hydrilla* also deviate from that of the typical angiosperm plastid genome. A large extension of the inverted repeats into the small single copy region in *Najas* results from the incorporation of the entire ndhH-ndhD operon, composed of the psaC gene, along with the pseudogenized remnants of the remaining ndh genes (ndhHψ, ndhEψ, ndhDψ). In *Hydrilla*, two inversions in the large single copy region are reported. Additionally, a number of genes are highly divergent in both *Najas* and *Hydrilla*.

Comparisons are made between 106 plastomes, from lineages with atypical and conservative evolution. Several patterns are evident in lineages with aberrant plastome evolution, including loss or divergence of a number of plastid genes (e.g., accD, clpP1, infA, ndh, and the PEP polymerase genes), along with perturbations in the PrrnP1 promoter region of the rrn operon.

Introduction

The plastid genome in the majority of angiosperms is generally regarded as circular molecule of double-stranded DNA restricted to between 108 kb and 160 kilobases (kb) in size (Palmer 1991), with conservation of gene structure, content and order characterizing most plants (Palmer 1991, Guisinger et al. 2011, Ruhlman & Jansen 2014). Typically, in most flowering plants, large and small single copy regions (LSC: ~80-90 kb, SSC: ~16-27 kb) are separated by two repeat regions, in reverse orientation (IR_A and IR_B: ~20-28 kb each) (Chumley et al. 2006).

While these generalities apply to the majority of angiosperm plastomes, aberrant plastome sizes, ranging from approximately 11 kb in the endoparasite *Pilostyles aethiopica*, Apodanthaceae (Bellot & Renner 2016) to 218 kb in *Pelargonium* ×*hortorum*, Geraniaceae (Chumley et al. 2006), have been reported. Likewise, anomalous lineages with convergent losses in otherwise highly conserved genes have also been discovered, for example, gnetophytes and pines (Werner et al. 2009), clades of Geraniaceae (Blazier et al. 2011), Ericaceae (Fajardo et al. 2013), Orchidaceae (Pan et al. 2012), and the aquatic order Alismatales, where loss/ pseudogenization of the 11 chloroplast ndh genes has been reported (Iles et al. 2013, Peredo et al. 2013, Ross et al. 2016). These genes encode subunits of the NADH dehydrogenase-like (NDH) complex associated with Photosystem I, and are implicated in optimizing the rate of cyclic electron flow and reducing photo-oxidative stress under field conditions (Martín et al. 2009).

Along with loss of the NDH complex, initial plastome sequencing of one member of the Alismatales, *Najas flexilis*, revealed three other unusual plastome features: including a novel intergenic spacer region within the psaA-psaB-rps14 operon, a region which is highly conserved across all land plants (Peredo et al. 2012, Peredo et al. 2013); loss of the rpl16 intron, which is

also widely conserved across angiosperms (Campagna & Downie 1998); and highly divergent rpo genes, encoding for the plastid-encoded polymerase (PEP) (Chapter 2).

Transcription of plastid genes is highly complicated, with at least three different polymerases implicated (Lerbs-Mache 2011), the plastid-encoded polymerase (PEP), and two nuclear-encoded polymerases. In most higher plants, the rrn operon is transcribed by the PEP polymerase (of cyanobacterial ancestry) from a σ^{70}-type promoter (PrrnP1); however, a second promoter (PrrnP2) is also recognized by a nuclear-encoded polymerase in tobacco; and in both spinach and tobacco, a third promoter (Pc) is present upstream of the rrn operon, and is probably also recognized by a nuclear-encoded polymerase (Suzuki et al. 2003).

The eubacterial-type σ^{70}-type promoter recognized by PEP typically contains a -35 (TTGACT) and -10 (TATATT) element. To investigate important promoter elements in the PrrnP1 promotor, Suzuki et al. (2003) performed promoter dissection, and identified an essential hexameric sequence (GTGGGA; the rRNA operon up-stream activator [RUA]), upstream of the -35 element. They proposed that sigma factor interactions with the -10 element in the PrrnP1 are replaced by direct PEP-RUA interaction (or by interactions between PEP and a RUA-binding transcription factor). Their analysis suggested that this RUA element is highly conserved across angiosperms.

Najas

The cosmopolitan aquatic genus *Najas* L. (Hydrocharitaceae) consists of approximately 40 species (Triest 1988), separated into two subgenera, *Najas* and *Caulinia* (Ascherson 1864, Rendle 1899). As mentioned, previous sequencing of *N. flexilis* from subgenus *Caulinia* revealed loss or pseudogenization of the plastid genes coding for the NDH complex; however, at that time

Sanger sequencing failed to confirm loss of these genes in subgenus *Najas*. As part of our systematic revision of North American *Najas*, complete plastomes from nine North American *Najas* taxa, representing both subgenera are characterized here. Additionally, a partial plastome assembly of another member of Hydrocharitaceae, *Hydrilla verticillata*, is provided. *Hydrilla* is an Old World species, which is recognized as one of the most invasive aquatic plants in North America (True-Meadows et al. 2016). To date, two North American introductions of *Hydrilla* have been documented by the presence of discrete genetic lineages: a dioecious (but female) biotype and a monoecious biotype (Cook & Lund 1982, Benoit 2011).

Some questions remaining to be resolved are 1) is the NDH complex missing in both subgenera of *Najas*, 2) is this complex also lost in *Hydrilla* and 3) are there other genes in *Najas* or *Hydrilla* that are divergent with respect to other angiosperms?

With increased Illumina reads to assist with the assembly of chloroplast genomes in this study, an assembly error in our previously reported plastome for *Najas flexilis* [NC_021936.1] was recognized; accordingly, amendments to the plastome size and location of the IR borders will be provided here.

Additionally, as initial investigations of the *Najas* and *Hydrilla* assemblies revealed that the PrrnP1 promoter region was divergent in both *Najas* and *Hydrilla*, with respect to that reported by Suzuki et al. (2003), a comparison of this region in 106 angiosperms, representing both conserved and atypical plastomes, was made to determine whether loss of this region is a shared trait amongst taxa with aberrant plastomes. It was also questioned whether there were other shared characteristics between atypical plastid genomes.

Materials and Methods

Taxon sampling

Sampling represented both subgenera, with two accessions from subgenus *Najas* (*N. major* and *N. marina*). Until recently (Rüegg et al. 2016), these two taxa had both been united under *N. marina* (Figure 1). From section *Caulinia*, two separate accessions of *N. minor* (USA1 and USA2), were included as part of another study (Les et al. 2015). These accessions represent different introductions of this invasive taxon into North America. These individuals were sequenced to determine sequence divergence at the chloroplast level and provide additional phylogenetic markers to trace introductions of this invasive taxon. The remaining taxa represent accessions from section *Americanae* and *Euvaginatae* in North America and were sequenced as part of our systematic study of North American *Najas*. A single accession of dioecious *Hydrilla* from Florida was also sequenced.

Plant specimens were preserved in NaCl/CTAB solution (Rogstad 1992) with a corresponding dried voucher specimen deposited in the CONN herbarium (Appendix A). Additional specimens were obtained from collaborators as dried plant material and processed similarly.

DNA isolation and sequencing

Total genomic DNA was extracted following Doyle & Doyle (1987). Molecular confirmation of the identity of each accession was performed by sequencing the nuclear ITS, and chloroplast rbcL and trnK/matK regions as described previously (Les & Tippery, 2010). Genomic DNA (500 ng) for five individuals was subjected to 454 library preparation and shotgun sequencing using the GS FLX Titanium pyrosequencing platform (454 Life Science Corporation, Branford, CT) at the Center for Applied Genetics and Technology, University of Connecticut.

Additionally, multiplex libraries for all ten accessions were prepared using the Nextera kit (Illumina, San Diego, CA). Library products were cleaned using Agencourt AMPure XP PCR purification systems (Bechman Coulter, Indianapolis, IN), quantified using a Qubit Quati-iT DNA Assay (Invitrogen, Carlsbad, CA) and checked for quality and to ensure a library size distribution of between 300 to 1200bp using a BioAnalyzer 2100 High Sensitivity DNA Chip (Agilent, Santa Clara, CA). Sequencing of 250bp paired-end reads was performed using a 500-cycle Illumina MiSeq V.2 sequencing kit (Illumina, San Diego, CA), at the Department of Molecular and Cell Biology, University of Connecticut.

Plastome assemblies

Following filtering and quality trimming (0.05 error probability), paired-end reads (combined with 454 reads when available) were *de novo* assembled using CLC Genomics Workbench v6.5 (CLC Bio, Aarhus, Denmark), using default assembly parameters. Plastid contigs were identified using BLAST searches against a custom database of plastid protein-coding genes for *Najas flexilis* [Genbank: NC_021936.1] in Geneious V6.1 (http://www.geneious.com, Kearse et al. 2012). Contigs were elongated by iteratively remapping reads until contigs overlapped. Independently, paired-reads were directly mapped to the *N. flexilis* chloroplast assembly. Concatenated consensus sequences were then verified by mapping reads and inspecting the resulting assemblies by eye for mismatches or unexpected drops in coverage. Assembly conflicts at the IR and SSC junctions between the newly sequenced plastids and the original *N. flexilis* plastome, along with high sequence divergence at this region amongst all taxa, required designing new primers to span the IR junctions in most taxa. Additionally, as initial mapping of *Hydrilla* assemblies to both *Elodea* and *Najas* suggested that inversions were present in regions of the *Hydrilla* plastome, de novo contigs, compiled from CLC assemblies,

were aligned to the Elodea chloroplast using MUMmer 3.0 (Kurtz et al. 2004), to evaluate these assemblies and determine minimum length matches and rearranged regions between *Hydrilla* and *Elodea*.

Primers were designed with Primer3 (Untergasser et al. 2012), as implemented in Geneious, to span junction regions, inverted regions, and for gaps or areas of low coverage in some taxa. PCRs were conducted in a 12.5 µl total reaction volume using 0.15 mM each dNTP (Promega, Madison, WI, USA), 0.4 µM each primer, 1x Titanium Taq® reaction buffer with 0.065 µl Titanium Taq® polymerase (Clontech, Mountain View, CA, USA), 1.0 M betaine (Affymetrix) and 20 ng of template DNA. Optimal annealing temperatures for each primer set were tested by gradient PCR (52^{o} C to 59^{o} C). PCR products were cleaned with ExoSAP-IT® (Affymetrix) and sequenced in both directions using BigDye® V1.1 (Life Technologies). Sequencing products were cleaned with SephadexTM G-50 columns (GE Healthcare Bio-sciences AB) and separated on an ABI PRISM 3130 genetic analyzer (Applied Biosystems).

Annotation

The consensus sequence for *N. major* was annotated in DOGMA (Dual Organellar GenoMe Annotator) (Wyman et al. 2004) with an identity cutoff for protein-coding genes of 35% and tRNA identity cutoff of 90% (e value: 1 e^{-5}), coupled with manual corrections for start and stop codons. Consensus sequences for all nine accessions, with one IR removed, were aligned using MAFFT V7.017 (Katoh et al. 2002) with the FFT-NS-i x1000 algorithm, scoring matrix 100PAM/k=2, gap open penalty 2.01, with an offset value of 0.123, with reference to *Najas flexilis* and *Elodea canadensis* [Genbank: NC_018541.1], to verify annotations and determine intron/exon positions. Conserved domains for open reading frames (ORFs) of divergent genes

were predicted using the Conserved Domain Database at NCBI (CDD v3.16), specifying an E-value of 0.01, with low-complexity filters applied (Marchler-Bauer et al. 2010). An annotated GenBank file was used to draw a circular plastid genome map of *N. major* using organellarGenomeDraw (OGDRAW) (Lohse 2013).

Repeat sequence analysis

REPuter (Kurtz et al. 2001) was employed to search for all forward (direct) and inverted (palindromic) repeat sequences of minimum 12 nt length (Hamming distance: 3) in the *Najas* plastomes. For the repeat analysis, one copy of the IR was removed to prevent duplication of counts. As REPuter overestimates the number of repeats within a given region, redundant repeats were discarded with only the longest repeat retained.

Simple sequence repeats (SSRs or microsatellites) were detected using Phobos v. 3.3.11 (Mayor, 2010) with thresholds of 8,5,4,3,3 and 3 repeat units for mono-, di-, tri-, tetra-, penta and hexanucleotide repeats respectively. Repeats were also evaluated for a number of other monocot taxa and *Nicotiana tabacum* to allow a comparison with *Najas* (Appendix A).

mVISTA

Sequence similarity among the nine *Najas* plastomes, was plotted using mVISTA (Frazer et al. 2004), using LAGAN global multiple alignment (Brudno et al. 2003) and a minimum percent conservation identity of 70% over a conservation window of 100bp. *Najas major* was used as the reference sequence. Percent similarity and patristic distances (HKY corrected) for entire plastomes were calculated in Geneious.

Plastid rrn operon

To investigate common features that might be shared by atypical plastomes, for example, loss of the RUA element (Suzuki et al. 2003), divergence in the accD, clpP, infA, ndh and rpo genes, along with large scale structural perturbations in the plastome, 106 angiosperm plastid genomes were downloaded from GenBank. Included in this sample were plastomes reported as having atypical plastome evolution in the literature, along with a random sample representing a range of angiosperm families. The rrn operon region was extracted from these plastomes and surveyed for the RUA , -35 and -10 elements. In addition, the accD and clpP genes were surveyed for conserved sites; and the presence/absence of the infA gene, and of the rpl16 intron in these plastomes was also recorded.

Results

The number of combined Illumina and Roche 454 assembled chloroplast reads ranged from between 24,081 in *Najas wrightiana* to 225,012 in *N. minor* (Appendix A). With increased Illumina read coverage in this study, amendments to the previously reported plastome sequence for *N. flexilis* [Genbank: NC021936] (Peredo et al. 2013) are reported here, as follows:

A single copy of the rps15 gene and ndhH-D operon (ndhHψ, ndhEψ, psaC, ndhDψ), residing in the SSC, was previously reported; however, this entire region is duplicated within the IRs, resulting in a highly reduced SSC region across *Najas* (2228 bp in *N. flexilis*) and a longer overall sequence length of 161,093 bp in *N. flexilis*.

In addition to the previously described non-canonical features of the *Najas* plastome, (psaA-psaB spacer, rpl16 intron loss, ndh gene loss/pseudogenization and elevated rates in the PEP polymerase genes); additionally, three genes (accD, clpP and infA) are reported here as having

highly divergent ORFs in both *Najas* and *Hydrilla*, and further research will be necessary to determine their functionality.

accD

The accD gene, encoding a subunit of acetyl-CoA carboxylase is highly divergent in *Najas* and *Hydrilla*. Maximum blast hits for the accD ORF results in 36% query coverage and 86% identity, with only ~580 bp at the 5' end identified as conserved (NCBI Conserved Domain). The conserved c-terminal motif II (Lee et al. 2004), hypothesized as the catalytic site is preserved; therefore, this gene may be functional, but highly divergent in these taxa. A number of repeat motifs are present in the 3' region resulting in length variation within *Najas*, and a repeat motifs are present at the point where both taxa diverge from other angiosperms. However, ORFs ranged in length from 1452-1467 bp in the *Najas* taxa sequenced here, which is typical of the length of this gene across angiosperms (Appendix B). *Hydrilla* however, has a long insertion of ~ 430 bp, relative to the other taxa here.

clpP1

The chloroplast clpP1 gene is highly conserved in flowering plants, and normally contains three exons, which code for a proteolytic subunit of an ATP-dependent protease (Erixon & Oxelman 2008). Exons one and three of the clpP1 gene in *Najas* have insertions of novel DNA, and Blastn searches result in no significant hits for these exons; however, insertions do not introduce frame shifts or early stop codons. Three amino acid residues (Ser-97, His-122 and Asp-171) have been identified in this gene as components of its catalytic triad in *E. coli* (Wang et al. 1997). In *Najas*, the aspartate residue is converted to a proline. *Hydrilla*, also contains insertions in exon three, but the catalytic triad is preserved. Mutation of this residue in bacterial ClpP1 orthologues

eliminates protease activity (Zeiler et al. 2013). Another unusual feature associated with this gene in both *Najas* and *Hydrilla* is that they lack the clpP NEP Type II -53 promotor, which is conserved in monocots, eudicots, conifers and liverworts (Sriraman et al. 1998, Liere et al. 2011). Appendix B contains an alignment of Najas and Hydrilla with other angiosperms at the exon three and the 5' intergenic region and shows the divergence of this promoter *region in these two taxa*. Williams et al. 2015 also reported the conversion of the third aspartate residue (to valine) in *Acacia lingulata*, along with ratios of nonsynonymous to synonymous substitutions (dn/ds) indicating relaxed selection on this gene in *Acacia*. However, RT-PCR transcripts of clpP1 in *Acacia* were readily detectible and both introns were correctly spliced out, indicating that the clpP1 protein might still be synthesized in the *Acacia* plastid.

infA

The infA gene, coding for translation initiation factor 1 is also highly divergent in *Najas*. Predicted ORFs in this region give no significant blast hit in GenBank and are not recognized as infA in a conserved domain search. The *Najas* ORFs contain three insertions relative to infA in other angiosperms, with a total gene length of 327 bp as opposed to 234 bp in *Elodea* and most other angiosperms (Appendix B). The start of the gene is dominated by a poly-A tract and a comparison of this gene between *Najas* sequenced here and the infA gene in *N. guadalupensis* (Ross et al. 2015) shows a 279 bp duplicated region at the beginning of the gene in *N. guadalupensis*. However, as no reference was made to assembly difficulties or divergence of the gene in that study, Sanger sequencing or longer reads over this region would be necessary to determine whether this is an assembly artifact in *N. guadalupensis*. Nuclear transfer of infA has been demonstrated in many rosids (Millen et al. 2001); however, no divergent reads were located in *Najas* assemblies, which might indicate functional nuclear copies. All insertions in *Najas*,

relative to other angiosperms, maintain codon frames and no stop codons are introduced, so functionality of this gene in the *Najas* chloroplast will need to be further determined.

On the other hand, the assembly of this gene in *Hydrilla* indicated that there were paralogous sequences in this region, and in fact the rpl16-rpoA operon (which normally houses this gene) could not be assembled with confidence in this study (discussed later).

General plastome characteristics of Najas

The nine *Najas* chloroplast genomes range in size from 156,791 bp in *Najas minor* [USA1] to 161,478 bp in *Najas wrightiana* (Figure 2), with LSC, SSC and IR regions of relatively similar lengths across taxa (Appendix A). Gene content and order are identical in all nine taxa, (a list of genes is provided in Peredo et al. 2013). Total plastome similarity ranged from 89.8% similarity between *N. major* and *N. flexilis* to 100% similarity (73 differences) between the two representatives of *N. minor* [USA1 and USA2]. Figure 3 shows pairwise similarity and patristic distances (HKY corrected). An mVISTA alignment of the nine plastomes (one IR removed) shows coding and non-coding regions with >70% similarity to *N. major* (Figure 4). Most differences in coding regions occur between the two subgenera, *Najas* and *Caulinia*. Although extensive chromosomal rearrangements have been demonstrated between *N. major* and *N. marina* (Viinikka 1976), until recently, these two taxa have been recognized as the same species (Rüegg et al. 2016). Both taxa are conserved at the coding level, with most differences occurring at the intergenic level, and regions within pseudogenized genes (e.g. ndhB and ndhE).

Earlier attempts to detect a presence of the 11 ndh genes within subgenus *Najas* (Peredo et al. 2013) were unfruitful due to lack of primer specificity. Complete loss of six genes (ndhA, F, G, I, J, and K) and pseudogenization of the remaining five genes of the complex (ndhB, C, D,

E and H) is reported here. Sequence similarity with *Elodea canadensis* of remaining ndh pseudogenes across *Najas* is given in Appendix B.

GC content is similar across all taxa, ranging from 27.3 - 29.3% in the SSC, 35.1 - 35.9% in the LSC region and between 41.8 - 42% in the IR regions (Appendix A). The relatively high GC content in the IRs is associated with the four genes of the rRNA operon, and the relatively low GC content in the SSC region reflects the fact that only two protein-coding genes (ccsA and rpl32), along with a single transfer RNA (tRNA-Leu), are present in the SSC in *Najas*, due to the large extension of the IR regions into the SSC region. The SSC region ranges in size from just 2228bp in *N. minor* [USA1] to 3391bp in *N. wrightiana*. Although the IR boundaries are dynamic, and movement of the IRs in and out of the SSC region occurs frequently (Plunkett & Downie 2000), these boundary shifts normally occur within the ycf1 gene, with large scale IR extensions being highly unusual. A comparison of the relative size of the SSC region between *Najas* and other sequenced alismatids and related monocots shows the large size reduction of this SSC region in *Najas* (Figure 5). The ycf1 and rps15 genes, along with the ndhH-D operon (containing psaC for Photosystem I) is duplicated in the IRs in all nine plastomes. Within the ndhD-H operon, *N. major* is lacking ndhE entirely, and only 185 nucleotides with sequence homology to ndhD is retained at the 5' end of the gene in *N. marina* (Appendix B).

The four junctions between the two single copy regions and the IRs are termed: J_{LB} (LSC/IR_B), J_{SB} (SSC/IR_B), J_{SA} (SSC/IR_A) and J_{LA} (LSC/IR_A). A comparison of these regions across the nine *Najas* plastomes is given in Figure 6. These boundary regions are dynamic even within these closely related taxa and contain highly AT-rich nucleotide regions. For example, in *N. major*, percentage AT composition corresponds to 86% (rps19-rpl2), 76% (ndhDψ -rpl32), 80% (ccsA-ndhD) and 85% (rpl2-trnH) at these boundaries.

In *N. major* and *N. marina* the J_{LB} junction occurs within the rps19 gene resulting in a duplication of 6 bases at the beginning of this gene in the IRs. This is similar to the arrangement in *Elodea canadensis* which has 9 bp of this gene duplicated.

Distances between the three remaining SSC genes (rpl32 - trnL - ccsA) are relatively conserved across taxa, with most SSC length variation occurring on either side of these genes. The rpL32 gene at J_{SB} is situated 135 bps from IR_B in *N.* marina and as far away as 1392 bp in *N. wrightiana*; however, in *N. flexilis* it is only 233 bp from the IR_B/SSC junction. At J_{SA} in *N. major*, four bp at the 3' end of ccsA is situated within IR_A, whereas in *N. marina*, 192 bp of the 3' end of ndhDψ is situated in the SSC. All other taxa have the entire ndhH-D operon in both IRs and the entire ccsA gene in the SSC.

The majority of monocots have a trnH-rps19 gene cluster duplicated in the IRs, however the J_{LB} and J_{LA} boundary regions are more dynamic within Alismatales, moving from within the rps19 gene, to between rps19-trnH and the trnH-rpl2 intergenic region (Wang et al. 2008). This study shows that even within *Najas*, these boundary regions are highly dynamic, with J_{LB} within the rps19 gene in *N. major* and *N. marina*, and in the rps19-trnH intergenic region in the remaining taxa.

Repeat analysis

Dispersed repeat analysis identified 27 to 50 direct and palindromic repeats of 30 bp or longer in the nine *Najas* plastomes (Figure 7). Apart from within ycf1 and ycf2, and a repeat shared between psaA and psaB, these repeat sequences were between tRNAs or intergenic regions. None of these repeats were located at the SSC/IR boundary regions. Several repeats were shared by all nine *Najas* plastomes. The number of dispersed repeats was similar to that of other

sequenced monocots, with the exception of the alismatid, *Sagittaria lichuanensis*, which had 100 dispersed repeats. It is notable that *Sagittaria lichuanensis* has the largest chloroplast reported amongst monocots (179,007 kb), which is second only in size to *Pelargonium* amongst angiosperms (Luo et al. 2016). *Sagittaria* also has a large 2.4 kb inversion in the LSC.

Interest in short tandem repeats (STRs) has gained with recognition that they may not simply be "junk DNA", and they are now implicated in the accelerated evolution of coding and regulatory sequences (reviewed in Gemayel et al. 2010). Strand slippage and mispairing during DNA replication and repair appears to be the regarded as the mechanism of mutation of these STRs (Fan & Chu 2007). In chloroplasts, with rearrangements or atypical length variation, repeat comparisons are often made, as correlations between such abnormalities and high levels of STRs have been observed (Palmer, 1991)

In *Najas*, the greatest number of tandem repeats comprised homopolymers (Appendix A). The number of homopolymers ≥ 7 bp in length ranged between 194 in *N. filifolia* to 212 in *N. marina* (≥ 8 bp in length ranged between 72 in *N. flexilis* to 91 in *N. major*), with A or T nucleotides exceeding G or C's by a ratio of ~ 14:1.

While many genes had poly A repeats, particularly at the 3' ends, which resulted in the genes being extended into the intergenic regions; other repeat motifs were primarily associated with intergenic regions (only mononucleotide repeats were found in introns), and were rarely located in coding regions (with the exception of accD, rbcL, rps4, ycf1 and ycf2). Indels between the *Najas* chloroplasts were most often associated with these repetitive sequences. However, *Najas* does not appear to have an elevated STR content in comparison to other sequenced alismatids and related monocots assessed here (Appendix A).

***Hydrilla* assembly**

A total of 67,639 reads were assembled into five plastid contigs for *Hydrilla*. The longest of these was a LSC contig, corresponding to the region from psbK to petD in *Elodea* (~80,300 bp). Within this region, two segmental inversions were observed in *Hydrilla*. These were confirmed by Sanger sequencing. One inversion occurs between trnV-UAC and the intergenic region between the ndhJ pseudogene and trnT-UGU (Figure 8). This results in the loss of exon two in trnV gene. The other inversion occurs between two intergenic region, the petA to psbJ, and accD to psabI regions (Figure 9).

As mentioned, the rpl16-rpoA operon (rpl16, pl14, rps8, infA, rpl36, rps11, rpoA), which normally occurs after the petD gene, could not be assembled with confidence with these reads. A number of AT rich repeats were associated with gaps in this contig, and all attempts to Sanger sequence over these gaps failed. Additionally, paralogues were recognized by the presence of snps at two loci associated with intergenic regions in this operon.

Similarly, two gaps remained in the IR contigs. Even though there was deep coverage over the ycf1 and ycf2 genes, and the rrn operon, these regions could not be assembled into longer contigs, and longer sequence reads will be necessary to confidently assign the locations of these contigs. Unusually, reads mapped to the *N. flexilis* or *Elodea* ccsA gene could not be extended past the 5’ end of this gene, no matter how the mapping algorithm was relaxed. Similarly, only ~ 300 bp could be built past the rpl32 gene. These are the two genes that border the SSC in *Najas* (with trnL in between, see Figure 6); however, these regions were not particularly AT rich in *Hydrilla*, and it is unclear why no reads could be assembled in this region.

Hydrilla also shows loss/pseudogenization of the 11 ndh genes, and the only fragments of these genes retained in *Hydrilla* are ndhB (exon 2), ndhJ and ndhK. These two later genes are located in one of the LSC inversion regions (with the third gene from this operon, ndhC missing).

Plastome comparisons

Of the 106 plastomes surveyed here, fifty possessed the RUA element (Suzuki et al. 2003), along with the -35 and -10 elements, at a conserved distance of 145 bases between the -35 element and the 5' end of the rrn16 gene (Appendix C). Of these plastomes, only two species lacked the accD gene (*Acorus americanus* and *Acorus calamus*), one species had a pseudogenized copy (*Arbutus unedo*), with eight species lacking complete conservation of the five motifs of Lee et al. (2004). Motif II is estimated to be the conserved catalytic site (Lee et al. 2004).

All species had the clpP1 gene in the conserved orientation, with only one species lacking the two introns in the clpP1 gene (*Arbutus unedo*). Less conservation was observed in the infA gene. This gene was missing from 14 taxa in this group, with a pseudogenized copy present in two species, and less than half of the species (22) had a conserved gene length of 234 bp. Additionally the rpl16 intron was present in all of these taxa. Loss of this intron has been recorded for several taxa within this group; however, when the original gene annotation was checked in these species, it was revealed that exon one had been annotated incorrectly in all cases (10 taxa), and that the intron was in fact still present.

Interestingly, the five surveyed plastomes from the grass family (Poaceae), all had a single mutation (TATACT) in the -10 element, and shared a conserved distance of 147 bp between the -35 element, and the start of the rrn16 gene. These diverse grass species also shared a conserved 69 bp insertion in the rrn23 gene relative to the other species surveyed here. The

accD gene is lost across all Poaceae (as are the ycf1 and ycf2 genes in the inverted repeats). Additionally, length variation in the infA gene was observed in these taxa (either 324 bp or 342 bp).

Previously reported non canonical plastomes (e.g., in lineages of Caryophyllaceae, Fabaceae, Geraniaceae, Oleaceae and Onagraceae) all lacked conservation in this RUA region (Suzuki et al. 2003). In instances where all three elements (RUA, -35, -10) retain sequence conservation, some perturbation of the conserved distance to the rrn16 gene was observed (Appendix C). For example, *Anacardium occidentale* had all three elements, but at a greater distance (152 bp).

Twelve species were missing the accD gene, and in this group only 13 species had an intact gene with the five conserved motifs (Lee et al. 2004). Two taxa have lost the clpP1 gene, and in a further seven species the gene is pseudogenized (and in most of these cases the gene is duplicated in the IRs, or the LSC). A further 20 species have lost introns in this gene (in these cases, the gene was observed to be in an inverse orientation to the normal orientation). In the majority of taxa in this group, the infA gene is either lost or pseudogenized; and while the gene appears functional in 17 taxa, all (except six species) deviate from the more normal 234 bp length conserved across angiosperms. This was also the case for *Najas* and *Hydrilla*.

Sixteen species were missing the rpl16 intron, and this was restricted to only three families, Caryophyllaceae, Geraniaceae and Hydrocharitaceae (*Najas*). Once again, five species had incorrect annotations for this gene. The ndh genes have also been lost or pseudogenized in some of these species, and highly divergent PEP polymerase genes have also been reported. However

time did not permit a full survey of these genes across these groups and this will require further investigation.

Discussion

Sequencing of plastid genomes has gained pace considerably in the last number of years, and as of December 2017, there are 1958 completely sequenced angiosperm plastomes deposited in GenBank (https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?opt=plastid&taxid=3398). As more completely sequenced plastomes become available, it remains to be seen whether more lineages will display exceptions to the general model of the chloroplast; and whether any trends will emerge amongst such groups (for example, Guisinger et al. 2011, Martínez-Alberola 2013, Blazier et al. 2016, Williams et al. 2015, Wang et al. 2017).

Further characterization of the *Najas* plastome, in this study, reveals additional traits which deviate from the typical angiosperm plastome. Large IR movements are rare in plastomes (Downie & Jansen 2015), with illegitimate recombination being regarded as the most plausible explanation (Blazier et al. 2016, Downie & Jansen 2015). And yet, anomalous chloroplasts (occurring in diverse lineages) have been reported, where large-scale IR expansions have occurred, for example 76 kb in *Pelargionium ×hortorum* (Palmer et al. 1987), 11.5 kb in species of *Berberis* and *Mahonia* (Kim & Jansen 1995), and 12 kb in *Nicotiana acuminata* (Goulding et al. 1996). In *Najas* the movement of the entire ndhH-ndhD operon from the SSC into the IRs, has resulted in a highly reduced SSC region, containing just 3 genes in *Najas*.

Studies have also shown that regions with a high repeat content or “poly A tracts” are associated with IR junction movement, and these low complexity regions are hypothesized to

facilitate recombination (Wang et al. 2008, Dugas et al. 2015). A model, beginning with a DNA double-strand break (DSB), followed by strand invasion from the inverted IR region, facilitated by repetitive regions, has been proposed to explain IR expansion in many angiosperm lineages (Wang et al. 2008). However, Kwon et al. (2010) used a transgenic homing intron and its endonuclease from a *Chlamydomonas* chloroplast to investigate double strand breakage in *Arabidopsis* (which has a deficit of repeated sequences >25 bp). They found that the chloroplast could repair DSBs using very little sequence homology, and with little DNA loss. This contrasted with their results for *Chlamydomonas* (relatively repeat rich) where single strand annealing occurred most frequently between direct repeats >30 bp.

DNA repair mechanisms are expected to play important roles in the conservation of chloroplast DNA sequences, given the high degree of photooxidative stress encountered in the chloroplast, but as yet relatively little is known about the mechanisms involved in DSB repair in chloroplasts, or the proteins involved in the various pathways of repair (discussed in Kwon et al. 2010), or whether different chloroplasts have evolved different abilities to repair DSBs without extensive homology.

In *Najas* the SSC-IR boundaries are extremely AT-rich, with long poly A tracts and AT repeat motifs (as is the ycf1 gene in *Najas*). Thus, DSB followed by recombination between these low complexity regions seems a plausible mechanism to explain the large IR expansion in *Najas* relative to *Elodea* (the closest completely sequenced relative), which has the typical angiosperm conformation. Overall repeat content in *Najas* however was not exceptionally high and was similar to that of other monocots surveyed here. On the other hand, *Hydrilla* was observed to have a high number of repeat rich regions, which also contributed to undermining assemblies in this study.

Two inversions were observed within the LSC region in *Hydrilla*. One of these inversions occurs in the rps4 to atpE region, spanning the region containing the ndhC to ndhJ operon, and resulting in loss of the second trnV-UAC exon. Inversions in the trnV-psbE region are also observed in *Cuscuta exaltata*, a parasitic plant (McNeal et al. 2007), along with *Circaeaster agrestis* (Circaeasteraceaee), both of which have experienced ndh gene loss (Sun et al. 2017)

Divergent genes in *Najas* and *Hydrilla*

The plastid accD gene encodes the B-carboxyl transferase subunit of acetyl-CoA carboxylase (ACCase) (Sasaki et al. 1997). This carboxylase catalyses the formation of malonyl-CoA from acetyl-CoA, and is estimated to be the regulatory enzyme of fatty acid synthesis (Ohlrogge & Browse, 1995). Knockout experiments have shown that this gene is essential in *Nicotiana* (Kode et al. 2005), and yet this gene has been lost independently in a number of different lineages, including Acoraceae (Goremykin et al. 2005), Campanulaceae (Haberle et al. 2008), Fabaceae (Magee et al. 2010), Geraniaceae (Guisinger et al. 2008) and Poaceae (Konishi & Sasaki 1994). In *Trachelium* (Campanulaceae) a chimeric nuclear gene (n-accD) of chloroplast origin was identified. This gene was smaller than the normal plastid gene, consisting only of a target peptide fused to the carboxylase domain of a plastid-like accD gene (Rousseau-Gueutin et al. 2013). This study also demonstrated that the product of the nuclear gene was imported into the plastid. Similarly, in *Silene* (Caryophyllaceae), where independent lineages have elevated rates in accD, clpP and other genes, it has recently been demonstrated that a duplicated cytosolic ACCase gene, which is smaller than the plastid gene, has an N-terminal extension that is strongly predicted as a plastid-targeting peptide (Rockenbach et al. 2016). Given that the accD gene in both *Najas* and *Hydrilla* retains the 3’ end with the carboxylase function, this gene is likely functional. However, given the divergence of even this portion of the gene (the 5’ end is unalignable with the accD

gene from other angiosperms here), the use of this gene in phylogenetic reconstruction is questionable.

Like the ACCase complex, the CLP protease complex contains a single plastid encoded subunit (Nishimura et al. 2014). Clp proteases function in removing denatured polypeptides from the chloroplast (reviewed in Clarke 1999), with nuclear-encoded ATP-dependent chaperones necessary for unfolding substrates for proteolysis by the catalytic components of the complex (Nishimura et al. 2015). The clpP1 gene is essential in both tobacco (Shikanai et al. 2001) and *Chlamydomonas reinhardtii* (Huang et al. 1994), and it is one of the few genes consistently conserved in mycoheterotrophic (Delannoy et al. 2011) or non-photosynthetic parasitic plants (Funk et al. 2007). But in rare cases, this gene is missing from the chloroplast (Guisinger et al. 2011, Fajardo et al. 2013, Martínez-Alberola 2013, Yang et al. 2013, Yang et al. 2014, Wang et al. 2016) (and see Appendix C), or exhibits elevated rates of sequence evolution (Erixon & Oxelman 2008, Sloan et al. 2014, Williams et al. 2015, Rochenbach et al. 2016). It is hard to conclude whether this gene is functional in *Najas* and *Hydrilla*, and further work is necessary to determine whether the gene is transcribed and spliced correctly in these taxa, as in *Acacia lingulata* (or whether the product of a nuclear gene now functions in the plastid).

Three initiation factors are known from eubacterial translation mechanisms and while translation initiation in organelles is believed to be similar to that in eubacteria, as yet, relatively little is actually known about how chloroplast mRNAs are recognized by ribosomes and how translation is regulated (reviewed in Zerges 2000). The plastid infA gene codes for translation initiation factor 1 which is a presumed homologue of IFI in *E. coli* (Daniell et al. 2016), mediating the coming together (along with two nuclear-encoded factors) of the mRNA, ribosome and initiator tRNA-Met to begin translation (Millen et al. 2001). Although infA is present in all

bryophyte and fern lineages, multiple independent losses of the gene have been recorded in angiosperms, including most rosids (Jansen et al. 2007, Magee et al. 2010, Millen et al. 2011). Of interest is the fact that loss of this gene is especially evident in lineages known for their atypical plastid genome evolution, such as those with extensive rearrangements or perturbation of the IRs (Wicke et al. 2011). As with accD and clpP1, the lack of premature stop codons suggests that this gene may also functional in *Najas* and *Hydrilla*, but existing in a highly divergent state compared with the majority of angiosperms.

The infA gene is located in the reverse transcribed rpoA-rpl16 operon, (rpl16, rps8, infA, rpl36, rps11, rpoA) located in the LSC. Another divergent feature of this operon in *Najas* is loss of the rpl16 intron. Campagna & Downie (1998) surveyed presence/absence of the rpl16 intron in 210 species from 86 angiosperm families, and found this intron to be highly stable component of the chloroplast genome of angiosperm plants, only absent in their survey from three families (Geraniaceae, Goodeniaceae and Plumbaginaceae). Loss of this intron is incorrectly reported in a number of species surveyed here, due to incorrect annotation of exon one of this gene (see appendix C), and this study found that only taxa from Caryophyllaceae (*Silene noctiflora*), Geraniaceae and *Najas* lacked this intron. In many cases in Geraniaceae, loss of this intron appears to be associated with perturbations to the rpl16-rpoA operon itself. As far as I am aware, no complete chloroplast sequences have been provided to date for the families, Goodeniaceae and Plumbaginaceae, and it will be of interest to see whether plastomes from these families are conserved or not.

Hydrilla, on the other hand, does contain the rpl16 intron; however, the rpl16-rpoA operon is also perturbed in *Hydrilla*. Mixed reads observed in assemblies of this region in *Hydrilla* may result from some form of heteroplasmy, or be due to duplicated regions within the

organelle itself, as, for example, is the case with the rpoA gene in *Pelargonium*, Geraniaceae (Blazier et al. 2016). Alternatively, there may be additional nuclear copies associated with genes that are normally present in this operon. Longer sequence reads will be necessary to assemble the *Hydrilla* plastome with more confidence.

Although convergent ndh gene loss has occurred in the Alismatales; within Hydrocharitaceae, the only clade represented by loss of these genes is the subfamily *Hydrilloideae*. Loss of these plastid genes in *Hydrilla* is confirmed here, and lends further support to the affinities of this group.

Plastome patterns

Some interesting patterns are evident in the plastomes compared here; however, finding explanations for these patterns is likely to be more problematic. It is certainly noteworthy that all of the well known lineages with atypical plastid genomes have some perturbation to the PrrnP1 promoter region. Additionally, it is conspicuous that within certain clades, for example, Oleaceae, those taxa with regular plastomes (*Forsythia suspensa* and *Olea europaea*) retain the RUA, -35 and-10 regions at a conserved distance from rrn16 (along with the accD, clpP1 and infA genes), whereas other species within that family, with irregular plastomes (*Jasminum nudiflorum* and *J. tortuosum*) lack conservation in this promoter region. These plastomes have also experienced accD gene loss, clpP1 intron loss, and have long poly C runs within the infA gene. Similarly, within the genus *Silene* (Caryophyllaceae), taxa that lack the PrrnP1 region (*S. chalcedonica*, *S. conica*, *S. noctiflora*) are missing genes or introns, and have large inversions in their plastomes. In fact *S. noctiflora* is the most complicated plastome within Caryophyllales

(Kang et al. 2017), yet the regular plastomes (*S. latifolia* and *S. vulgaris*) retain the normal gene features.

Although chloroplast transmission is believed to be principally uniparental and maternal, a number of plant groups exhibit paternal (e.g., pines) or biparental inheritance, for example, some species of Campanulaceae, Fabaceae, Geraniaceae and Onagraceae (Corriveau & Coleman 1988), and it has been observed that most lineages displaying atypical displaying atypical plastome organization inherit their plastomes biparentally (Ruhlman et al. 2017). No such evidence exists for biparental inheritance prevailing in either *Najas* or *Hydrilla*, but of course this possibility cannot be ruled out.

Additionally, the extent to which different conformations of the plastid genome may reside in the cell, and how this might play into a propensity for structural rearrangements through intermolecular recombination is just beginning to be researched (Ruhlman et al. 2017).

Conclusion

In summary, *Najas* plastomes deviate from the typical angiosperm plastome in the following aspects:

- Loss/pseudogenization of the 11 ndh genes.
- Highly reduced SSC region, consequent with the incorporation of the entire ndhH-ndhD. operon along with the rps15 and ycf1 genes from the IR region.
- Loss of the rpl16 intron in the rpoA - rpl16 operon.
- Highly divergent ORFs of accD clpP1 and infA.
- An unusual intergenic spacer in the highly conserved psaA/psaB operon.

- Elevated dn/ds ratios in two of the four rpo genes, encoding the PEP polymerase.

Although a circle could not be completed for *Hydrilla*, this taxon also shows unusually divergent genes along with loss of conserved genes and rearranged regions. These features place *Najas* and *Hydrilla* within a small group of divergent angiosperm lineages displaying atypical chloroplast evolution.

As more complete chloroplasts are sequenced, and more unusual plastomes are discovered, it will be of interest to see what these unconventional plastomes share in common, and what features set them apart.

References

Adam, Z. Adamska, I. Nakabayashi, K. Ostersetzer, O. Haussuhl, K. Manuell, A. Zheng, B. Vallon, O. Rodermel, S.R. Shinozaki, K. & Clarke, A.K. 2001. Chloroplast and mitochondrial proteases inArabidopsis. A proposed nomenclature.Plant Physiol.125, 1912 – 1918.

Bellot, S. & S.S Renner. 2016. The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. Genome Biol. Evol. 8, 189–201.

Benoit, L.K. 2011. Cryptic speciation, genetic diversity and herbicide resistance in the invasive aquatic plant *Hydrilla verticillata* (L.f.) Royle (Hydrocharitaceae). Ph.D. dissertation, University of Connecticut, Storrs, CT. Website (http://digitalcommons.uconn.edu/dissertations/AAI3492149).

Blazier, J.C. Guisinger, M.M. & R.K. Jansen. 2011. Recent loss of plastid-encoded ndh genes within *Erodium* (Geraniaceae). Plant Mol Biol. 76:263–272.

Blazier, J.C. Jansen, R.K. Mower, J.P. Govindu, M. Zhang, J. Weng M.L. & T.A. Ruhlman. 2016. Variable presence of the inverted repeat and plastome stability in *Erodium*. Annals of Botany.7: 1209-1220.

Bock, R. 2007. Structure, function, and inheritance of plastid genomes. In: Bock R (Ed) Cell and molecular biology of plastids. Springer, Berlin, pp 29–63.

Bofkin, L. & N. Goldman. 2007. Variation in evolutionary processes at different codon positions, Mol Biol Evol. 24: 513-521.

Braun A. 1864. Revision of the genus *Najas* of Linnaeus. Seemann's Journal of Botany 2: 274-279.

Brudno, M.S. Malde, A. Poliakov, Chuong, B. Do, Couronne, O. Dubchak I. & S. Batzoglou. 2003. Glocal alignment: Finding rearrangements during alignment. Bioinformatics 19: i54-i62. doi: 10.1093/bioinformatics/btg1005

Campagna, M.L. & S. R. Downie 1998. The intron in chloroplast gene rpl16 is missing from the flowering plant families Geraniaceae, Goodeniaceae, and Plumbaginaceae. Transactions of the Illinois State Academy of Science. 91: 1-11.

Clarke, A.K. 1999. ATP-dependent Clp proteases in photosynthetic organisms—A cut above the rest! Ann. Bot. 83: 593–599.

Cook, C.D.K. Gut, B. J. Rix, E. Schneller M. J. & M. Seitz. 1974. Water Plants of the World, a Manual for the Identification of the Genera of Freshwater Macrophytes. W. Junk, The Hague.

Cook, C.D.K. & R. Luond. 1982. A revision of the genus *Hydrilla* (Hydrocharitaceae).
Aquatic Botany 13:485-504.

Cook, C.D.K. 1982. Pollination mechanisms in the Hydrocharitaceae. In J.J. Symoens, S. S. Hooper and P. Compère [eds.], Studies on aquatic vascular plants, 1–15 Brussels: Royal Botanical Society of Belgium.

Corriveau, J.L. & A.W. Coleman. 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. Amer. J. Bot. 75: 1443-1458.

Davis, J.I. McNeal, J.R. Barrett, C.F. et al. 2013. Contrasting patterns of support among plastid genes and genomes for major clades of the monocotyledons. In:Wilkin P, Mayo SJ, eds. Early events in monocot evolution. Systematics Association Special Volume Series. Cambridge: Cambridge University Press, 315-349.

Darling, A.E.B. Mau and N.T. Perna. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS ONE* 5: e11147. doi:10.1371/journal.pone.0011147

Delannoy, E. Fujii, S. Colas des Francs-Small, C. Brundrett, M. & I. Small. 2011. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. Mol Biol Evol, 28:2077–2086.

Doyle, J.J. & J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19: 11–15.

Dugas, D.V. Hernandez, D. Koenen, E.J. Schwarz, E. Straub, S. Hughes, C.E. et al. 2015. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in clpP. Scientific Reports. 5:16958–16972.

Erixon, P. & B. Oxelman. 2008. Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast clpP1 gene. PLoS One. (1):e1386.
Fan, H, Chu, J.Y. 2007. A brief review of short tandem repeat mutation. Genomics Proteomics Bioinformatics 5:7–14.

Felsenstein, J. 1985. Confidence limits on phylogenies – An approach using the bootstrap. Evolution 39: 783–791.

Frandsen, P.B.B. Calcott, C. Mayer & R. Lanfear. 2015. Automatic selection of partitioning schemes for phylogenetic analyses using iterative k-means clustering of site rates. BMC Evolutionary Biology 15: 1–17.

Frazer, K.A. Pachter, L. Poliakov, A. Rubin, E.M. & I. Dubchak I. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 1; 32.

Funk, H.T. Berg S. Krupinska K. Maier U.G. & K. Krause. 2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. BMC Plant Biol. 7:45.

Gemayel, R. Vinces, M.D. Legendre, M. & Verstrepen, K.J. 2010. Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu. Rev. Genet. 44: 445–477.

Goremykin, V.V. Holland, B. Hirsch-Ernst, K.I. & F.H. Hellwig. 2005 Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. Mol Biol Evol. 22:1813-22.

Goulding, S.E. Olmstead, R.G. Morden, C.W. & K.H. Wolfe. 1996. Ebb and flow of the chloroplast inverted repeat. Mol. Gen. Genet. 252: 195–206.

Greiner, S. Wang, X. Rauwolf, U. Silber, M.V. Mayer, K. Meurer, J. Haberer G. & R.G Herrmann. 2008 The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera* : I. Sequence evaluation and plastome evolution. Nucl Acids Res 36:2366–2378.

Guisinger, M.M. Kuehl, J.V. Boore, J.L. & R.K. Jansen. 2011. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol 28: 583–600.

Haberle, R.C. Fourcade, H.M. Boore, J.L. Jansen, R.K. 2008 Extensive rearrangements in the chloroplast genome of Trachelium caeruleum are associated with repeats and tRNA genes. J Mol Evol 66: 350–361.

Hirose, T. & M Sugiura. 1997. Both RNA editing and RNA cleavage are required for translation of tobacco chloroplast ndhD mRNA: a possible regulatory mechanism for the expression of a chloroplast operon consisting of functionally unrelated genes EMBO J. 16: 6804-6811.

Hoang, D. T. Chernomor, O. von Haeseler, A. Minh, B.Q. & L.S. Vinh. 2017. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. doi: 10.1093/molbev/msx281.

Huang, C. Wang, S. Chen, L. Lemieux, C. Otis, C. Turmel, M. et al. 1994. The *Chlamydomonas* chloroplast *clpP* gene contains translated large insertion sequences and is essential for cell growth. Mol Gen Genet. 244:151–159.

Iles, W.J. Smith, S.Y. & S.W. Graham. 2013. A well-supported phylogenetic framework for the monocot order Alismatales reveals multiple losses of the plastid NADH dehydrogenase complex and a strong long-branch effect. *In* Wilkin, P. Mayo, S.J. [eds.], Early Events in Monocot Evolution, 1–28. Cambridge University Press Systematics Association Special Series, Cambridge, UK.

Ito, Y. Tanaka, N. Gale, S. Yano, O. & Li, J. 2017. Phylogeny of *Najas* (Hydrocharitaceae) revisited: Implications for systematics and evolution. Taxon, 66: 309–323.

Jansen, R.K. Cai, Z. Raubeson, L.A. Daniell, H. dePamphilis, C.W. Leebens-Mack, J.H. Müller, K.F. Guisinger-Bellian, M. Haberle, R.C. Hansen, A.K. Chumley, T.W. Lee, S.B. Peery, R. McNeal, J.R. Kuehl, J.V. Boore, J.L. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci USA 104:19369–19374

Katoh. K. Misawa, K. & K. Kuma & T. Miyata. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, Nucleic Acids Research. *30*: 3059–3066.

Kearse, M.R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649.

Kelchner, S.A. & L.G. Clark. 1997. Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). Mol. Phylogenet. Evol. 8: 385-397.

Knox, E.B. Downie, S.R. & J.D. Palmer. 1993. Chloroplast genome rearrangements and the evolution of giant lobelias from herbaceous ancestors, Mol Biol Evol. 10: 414-430.

Kode, V. Mudd, E.A. Iamtham, S. Day, A. 2005. The tobacco plastid *acc*D gene is essential and is required for leaf development. Plant J 44:237–244.

Kolaczkowski, B. Thornton, J.W. 2008. A mixed branch length model of heterotachy improves phylogenetic accuracy. Molecular Biology and Evolution. 25: 1054-1066.

Konishi, T. & Y. Sasaki. 1994. Compartmentalization of two forms of acetyl-CoA carboxylase in plants and the origin of their tolerance toward herbicides. Proc Natl Acad Sci USA. 91: 3598–3601.

Kuroda, H. & Maliga, P. 2003. The plastid *clpP1* protease gene is essential for plant development. Nature. 425:86–89.

Kurtz, S. J. V Choudhuri, E. Ohlebusch, C. Schleiermacher, Stoye J. & R. Giegerich. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic acids research 29: 4633–42.

Kurtz, S. Phillippy, A. Delcher, A.L. Smoot, M. Shumway, M. Antonescu, C. & S.L. Salzberg. 2004. Versatile and open software for comparing large genomes. Genome Biol.5, R12

Kwon, T. Huq, E. & D.L. Herrin. 2010. Microhomology-mediated and nonhomologous repair of a double-strand break in the chloroplast genome of Arabidopsis. Proc. Natl. Acad. Sci. USA. 107: 13954–13959.

Lee, S.S. Jeong, W.J. Bae, J.M. Bang, J.W. Liu, J.R. & C.D. Harn. 2004. Characterization of the plastid-encoded carboxyltransferase subunit (accD) gene of potato. Mol Cells 17: 422–429

Lerbs-Mache S. 2011. Function of plastid sigma factors in higher plants: regulation of gene expression or just preservation of constitutive transcription? Plant Mol Biol 76: 235–249.

Les, D. H. 1988. Breeding systems, population structure, and evolution in hydrophilous angiosperms, 75:819–835.

Les, D.H. Garvin, D.K. and Wimpee, C.F. 1993. Phylogenetic studies in the monocot subclass Alismatidae: evidence for a reappraisal of the aquatic order Najadales. Molecular Phylogenetics and Evolution, 2, 304–314.

Les, D.H. M.L. Moody & C.L. Soros. 2006. A reappraisal of the phylogenetic relationship in the monocotyledon family Hydrocharitaceae (Alismatidae). Aliso 22: 211–230.

Les, D.H. S.P. Sheldon & N.P. Tippery. 2010. Hybridization in Hydrophiles: Natural Interspecific Hybrids in *Najas* (Hydrocharitaceae). Systematic Botany 35: 736–744.

Les, D.H. & N.P. Tippery. 2013. In time and with water... the systematics of alismatid monocotyledons. *In* Wilkin, P. Mayo, S.J. [eds.], Early Events in Monocot Evolution, 118–164. Cambridge University Press Systematics Association Special Series, Cambridge, UK.

Les, D.H. E.L. Peredo, N.P. Tippery, L.K. Benoit, H. Razifard, U.M. King, H.R. Na, et al. 2015. *Najas minor* (Hydrocharitaceae) in North America: A reappraisal. Aquatic Botany 126: 60–72.

Liere K. Weihe A. & T. Borner. 2011. The transcription machineries of plant mitochondria and

chloroplasts: Composition, function, and regulation. J Plant Physiol 168: 1345-1360.

Lohse M. Drechsel O. Kahlau S. Bock R. 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 41: 575 – 581

Lopez P. Casane, D. Philippe H. Heterotachy, an important process in protein evolution. Mol Biol Evol. 2002;19:1–7.

Lowe, T.M. & S.R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955–964.

Luo, Y. Ma, P.F. Li, H.T. Yang, J.B. Wang, H. & D. Z. Li. 2016. Plastid phylogenomic analyses resolve Tofieldiaceae as the root of the early diverging monocot order Alismatales. Genome Biology and Evolution 8: 932 – 945.

Ma, J. Yang, B. Zhu, W. Sun, L. Tian, J. & X. Wang. 2013. The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. Gene. 528,120–131.

Marchler-Bauer, A. et al. 2010. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39: D225–D229.

Magee, A.M. Aspinall, S. Rice, D.W. Cusack, B.P. Sémon, M. Perry, A.S. Stefanović, S. Milbourne, D. Barth, S. Palmer, J.D. et al. 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Res 20: 1700–1710.

Maliga, P. Liere, K. Sriraman, P. & Z. Svab. 1999. A transgenic approach to characterize the plastid transcription machinery in higher plants. In: Argyroudi-Akoyunoglou JH and Senger H (eds) The Chloroplast: From Molecular Biology to Biotechnology, pp 317–323. Kluwer Academic Publishers, Dordrecht.

Martín, M. Funk, H.T. Serrot, P.H. Poltnigg, P. & B. Sabater. 2009 Functional characterization of the thylakoid Ndh complex phosphorylation by site-directed mutations in the ndhF gene. Biochim et Biophys Acta. 2009: 920–928. doi:10.1016/j.bbabio.2009.03.001.

Martínez-Alberola, F. del Campo, E.M. Lázaro-Gimeno, D. Mezquita-Claramonte, S. Molins, A. Mateu-Andrés, I. et al.: Balanced gene losses, duplications and intensive rearrangements led to an unusual regularly sized genome in *Arbutus unedo* chloroplasts. PLoS ONE 2013, 8:e79685. pmid:24260278

McNeal J.R., Kuehl J., Boore J. & C. dePamphilis. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. BMC Plant Biol 7:57

Michel, A. Arias, R.S. Scheffler, B.E. Duke, S.O. Netherland, M. & F. E. Dayan. 2004. Somatic mutation-mediated evolution of herbicide resistance in the nonindigenous invasive plant hydrilla (*Hydrilla verticillata*). Mol. Ecol. 13:3229-3237.

Nakkaew, A. Chotigeat, W. Eksomtramage, T. & A. Phongdara. 2008. Cloning and expression of a plastid-encoded subunit, beta-carboxyltransferase gene (*acc*D) and a nuclear-encoded subunit biotin carboxylase of acetyl-CoA carboxylase from oil palm (*Elaeis guineensis* Jacq.). Plant Sci 175:497–504.

Nishimura, K. & K.J. van Wijk. 2015. Organization, function and substrates of the essential Clp protease system in plastids. Biochim Biophys Acta. 9: 915-30.

Nishimura, K. Apitz, J. Friso, G. Kim, J. Ponnala, L. et al. 2015. Discovery of a unique Clp component, ClpF, in chloroplasts: a proposed binary ClpF-ClpS1 adaptor complex functions in substrate recognition and delivery. Plant Cell 27: 2677–2691.

Magee, A.M. Aspinall, S. Rice, D.W. Cusack, B.P. Sémon, M. Perry, A.S. et al. 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Research. 20:1700–1710.

Mayer, C. 2006–2010 Phobos 3.3.11. Available at http://www.rub.de/spezzoo/cm/cm_ phobos.htm

Millen, R.S. Olmstead, R.G. Adams, K.L. Palmer, J.D. Lao, N.T. et al. 2001. Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. The Plant Cell. 13: 645-658.

Minh, B.Q. Nguyen, M.A.T. & A. von Haeseler. 2013. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 30:1188-95.

Nguyen, L.T. Schmidt, H.A. vonn Haeseler A. & B.Q. Minh. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol Biol Evol, 32:268-274.

Ohlrogge, J. and Browse, J. 1995. Lipid biosynthesis. Plant Cell, 7, 957–970.

Palmer J.D. Osorio, B. & W.F. Thompson. 1988. Evolutionary significance of inversions in legume chloroplast DNAs. Current Genetics 14: 65–74.

Palmer, J.D. Osorio, B. Aldrich, J. & W.F. Thompson. 1987. Chloroplast DNA evolution among legumes:loss of a large inverted repeat occurred prior to other sequence rearrangements. Current Genetics 11: 275–286.

Palmer, J.D. 1991. Plastid chromosomes: structure and evolution. *In* I.K. Vasil [ed.].The Molecular Biology of Plastids, Cell culture and somatic cell genetics of plants, Vol. 7, 5-53. Academic Press, San Diego, California.

Pan, I.C. Liao, D.C. Wu, F.H. Daniell, H. Singh, N.D. et al. 2012. Complete chloroplast genome sequence of an orchid model plant candidate: Erycina pusilla apply in tropical *Oncidium* breeding. PloS one 7: e34738. doi:10.1371/ journal.pone.0034738.

Peltier, J.B. Ytterberg, J. Liberles, D.A. Roepstorff, P. & K.J. van Wijk. 2001. Identification of a 350-kDa ClpP protease complex with 10 different Clp isoforms in chloroplasts of *Arabidopsis thaliana*. J. Biol. Chem. 276: 16318–16327.

Peltier J.B. Ripoll D.R. Friso G. Rudella A. Cai Y. et al. 2004. Clp protease complexes from photosynthetic and non-photosynthetic plastids and mitochondria of plants, their predicted three-dimensional structures, and functional implications. J. Biol. Chem. 279: 4768–4781.

Peredo, E.L. D.H. Les, U.M. King & L.K. Benoit. 2012. Extreme conservation of the psaA/psaB intercistronic spacer reveals a translational motif coincident with the evolution of land plants. Journal of molecular evolution. 75: 184–97.

Peredo, E.L. King U.M. & D.H. Les. 2013. The Plastid Genome of *Najas flexilis*: Adaptation to Submersed Environments Is Accompanied by the Complete Loss of the NDH Complex in an Aquatic Angiosperm. *PloS ONE* 8. doi: 10.1371/journal.pone.0068591

Petersen, G. Seberg, O. Davis, J.I. & D. W. Stevenson. 2006. RNA editing and phylogenetic reconstruction in two monocot mitochondrial genes. Taxon, 55, 871–886.

Petersen, G. Seberg, O. Cuenca, A. Stevenson, D.W. Thadeo, M. Davis, J.I. Graham S. & T.G. Ross. 2016. Phylogeny of the Alismatales (Monocotyledons) and the relationship of *Acorus* (Acorales ?). Cladistics 32: 141–159.

Petersen, G. Cuenca, A. Zervas, A. Ross, G. T. Graham, S. W. Barrett, C. F. … Seberg, O. (2017). Mitochondrial genome evolution in Alismatales: Size reduction and extensive loss of ribosomal protein genes. *PLoS ONE*, *12*(5), 1–21. https://doi.org/10.1371/journal.pone.0177606

Rogstad, S.H. 1992. Saturated NaCl-CTAB solution as a means of field preservation of leaves for DNA analyses. Taxon 41: 701–708.

Raubeson, L.A. & R.K. Jansen. 2005. Chloroplast genomes of plants. *In* R. J. Henry, [ed.], Plant diversity and evolution: genotypic and phenotypic variation in higher plants, 45–68. CAB International, Wallingford, UK, pp 45 – 68.

Rendle, A.B. 1899. A systematic revision of the genus *Najas*. Transactions of the Linnean Society of London: Botany 5: 379–436.

Rockenbach, K.D. Havird, J.C. Monroe, J.G. et al. 2016. Positive selection in rapidly evolving plastid-nuclear enzyme complexes. Genetics, 204, 1507–1522.

Ronquist, F. Teslenko, M. Van Der Mark, P. Ayres, D.L. Darling, A. Höhna, S. Larget, B. et al. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539–542.

Rosendahl, C.O. & Butters, F.K. 1935. The genus *Najas* in Minnesota. Rhodora 37: 345– 348.

Ross, T.G. Barrett, C.F. Soto, M. Lam, V.K.Y. Henriquez, C.L. Les, D.H. Davis, J.I. et al. 2016. Plastid phylogenomics and molecular evolution of Alismatales. Cladistics 32: 160–178.

Rozen S & H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. *In*: S. Krawetz and S. Misener [eds.], Bioinformatics methods and protocols: Methods in molecular biology, 365–386. Humana Press, Totowa, New Jersey.

Rüegg, S. U. Raeder, A. Melzer, G. Heubl, & C. Bräuchler. 2016. Hybridisation and cryptic invasion in Najas marina L. (Hydrocharitaceae). Hydrobiologia.

Ruhlman, T.A. Zhang, J. Blazier, J.C. Sabir, J.S.M. & R.K. Jansen. 2017. Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. Am J Bot . 104:1–14.

Sasaki, Y. Kozaki, A. & M. Hatano. 1997. Link between light and fatty acid synthesis: Thioredoxin-linked reductive activation of plastidic acetyl-CoA carboxylase. Proc Natl Acad Sci USA 94: 11096–11101.

Song, Y. Yu, W.B. Tan, Y.H. Liu, B. Yao, X. Jin, J.J. Padmanaba, M. Yang, J.B. Corlett, R.T. 2016. Evolutionary comparisons of the chloroplast genome in Lauraceae and insights into loss events in the Magnoliids.Genome Biol Evol 9 (9):2354-2364.

Schattner, P. Brooks A.N. & T.M. Lowe. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Research 33: 686–689.

Shikanai, T. Shimizu, K. Ueda, K. Nishimura, Y. Kuroiwa, T. Hashimoto, T. 2001. The chloroplast *clpP* gene, encoding a proteolytic subunit of ATP-dependent protease, is indispensable for chloroplast development in tobacco. Plant Cell Physiol. 42:264–273.

Shinozaki, K. Ohme, M. Tanaka, M. Wakasugi, T. Hayashida, N. Matsubayashi, T. et al. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. 5: 2043-2049.

Sloan, D.B. Triant, D.A. Wu, M. & D.R. Taylor. 2014. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. Mol. Biol. Evol. 31:673–682.

Sloan, D.B. Triant, D.A. Forrester, N.J. Bergner, L.M. Wu, M. Taylor, D.R. 2014. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). Molecular Phylogenetics and Evolution. 72:82 – 89.

Sriraman, P. Silhavy, D. & P. Maliga. 1998. The phage-type PclpP-53 plastid promoter comprises sequences downstream of the transcription initiation site. Nucleic Acids Res. 26: 4874–4879.

Sugiura, M. 1989. The chloroplast chromosomes in land plants. Annu Rev Cell Biol. 5:51-70.

Sun YX, Moore MJ, Meng AP, Soltis PS, Soltis DE, Li JQ, Wang HC 2013: Complete plastid genome sequencing of Trochodendraceae reveals a significant expansion of the inverted repeat and suggests a Paleogene divergence between the two extant species. PLoS One 8(4):e60429.

Sun, Y. Moore, M.J. Lin, N. Adelalu, K.F. Meng, A. Jian S. & L. Yang 2017. Complete plastome sequencing of both living species of Circaeasteraceae (Ranunculales) reveals unusual rearrangements and the loss of the *ndh* gene family. BMC Genomics. 18:592.

Suzuki, J.Y., Sriraman, P., Svab, Z. & P. Maliga. 2003. Unique architecture of the plastid ribosomal RNA operon promoter recognized by the multisubunit RNA polymerase in tobacco and other higher plants. Plant Cell 15**:** 195–205.

Tonti-Filippini, J. Nevill, P.G. Dixon, K. & I. Small. 2017. What can we do with 1000 plastid genomes? Plant J. 4: 808-818.

Triest, L. 1988. A revision of the genus *Najas* L. (Najadaceae) in the Old World. Mémoires de l'Académie Royale des Sciences d'Outre-Mer . Classe des Sciences Naturelles et Médicales 22: 1 – 172.

True-Meadows, S. Haug, E.J. & R.J Richardson. 2016. Monoecious hydrilla — A review of the literature. J Aquat Plant Manag 54:1–11.

Untergasser, A. Cutcutache, I. Koressaar, T. Ye, J. Faircloth, B.C. Remm, M. & S. G. Rozen. 2012 Primer3 – new capabilities and interfaces. Nucleic Acids Research 40: 15.

Viinikka, Y. 1976. *Najas marina* L. (Najadaceae). Karyotypes, cultivation and morphological variation. Annales Botanici Fennici. 13: 119-131.

Wakasugi, T. Tsudzuki, T. & Sugiura, M. 2001. The genomics of land plant chloroplasts: gene content and alteration of genomic information by RNA editing. Photosynth. Res. 70: 107–118.

Wang, W.C. Chen, S.Y. Zhang, X.Z. 2016. Chloroplast genome evolution in Actinidiaceae: clpPloss, heterogenous divergence and phylogenomic practice. PLoS One 11:e0162324

Wang, R. Cheng, C. Chang, C. Wu, C. Su, T. & S. Chaw. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol Biol 8:36.

Wang, J. Hartling, J.A. & J.M. Flanagan. 1997. The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. Cell 91: 447–456.

Wang, Y.H. Qu, X.J. Chen, S.Y. & T.S. Yi. 2017. Plastomes of Mimosoideae: structural and size variation, sequence divergence, and phylogenetic implication. Tree Genetics & Genomes. 13: 41.

Weihe, A. Apitz, J. Pohlheim, F. Salinas-Hartwig, A. & Börner,T. 2009. Biparental inheritance of plastidial and mitochondrial DNA and hybrid variegation in Pelargonium. Mol. Genet. Genomics. 282:587–593

Willdenow, C.L. 1798. Détermination d'un nouveau genre de plante aquatique, nommé Caulinia, et observations générales sur les plantes aquatiques. Mém. Acad. Roy. Sci. Hist. Berlin. 1:78–90.
Williams, A.V. Boykin, L.M. Howell, K.A. Nevill, P.G. & I. Small. 2015. The complete sequence of the Acacia ligulata chloroplast genome reveals a highly divergent clpP1 gene. PLoS One 10: e0125768.

Wyman, S.K. Jansen R.K. & J.L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20: 3252–5.

Xia, X. 2017. DAMBE6: New tools for microbial genomics, phylogenetics and molecular evolution. J. Hered. 108: 431–437.

Xie, W. Lewis, P.O. Fan, Y. Kuo, L. & M.H. Chen. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection, Syst. Biol. 60: 150-160.

Yan L, Lai X, Li X, Wei C, Tan X, Zhang Y. Analyses of the complete genome and gene expression of chloroplast of sweet potato [*Ipomoea batata*]. PLoS One. 2015;10(4), e0124083.

Yang, Z. Wong, W.S. & R. Nielsen. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Molecular Biology and Evolution 22: 1107–1118.

Yang, J.B. Li, D.Z. Li, H.T. 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. Molecular Ecology Resources. 14:1024–31.

Yukawa, M. Tsudzuki, T. & M. Sugiura. 2005. The 2005 version of the chloroplast DNA sequence from tobacco (*Nicotiana tabacum*). Plant Molecular Biology Reporter. 23: 359–365.

Zeiler, E. List, A. Alte, F. Gersch, M. Wachtel, R. Poreba, M. et al. 2013. Structural and functional insights into caseinolytic proteases reveal an unprecedented regulation principle of their catalytic triad. Proc Natl Acad Sci USA. 110:11302 – 11307.

Zerges, W. 2000. Translation in chloroplasts. Biochimie 82:583–601

Figure Legends

Figure 1. Cladogram representing North American *Najas* taxa with plastomes sequenced in this study. Two accessions of N. minor [USA1 and USA2] are included, representing two separate introductions of this taxon into North America. These accessions were sequenced to contribute to another study (Les et al. 2015).

Figure 2. The chloroplast genome of *Najas major*. Genes inside the circle are transcribed clockwise, while genes outside are transcribed counter-clockwise. The dark grey inner circle corresponds to the GC content and the light-grey circle to the AT content.

Figure 3. Percentage pairwise identity (upper) and patristic distances (HKY corrected) (lower) between nine *Najas* plastomes.

Figure 4. mVISTA sequence identity plots comparing nine *Najas* chloroplast genomes with *N. major* as reference. Vertical scale represents % identity ranging from 50 to 100%. Coding and non-coding regions are marked in purple and pink, with ndh genes and rRNA marked in brown and blue respectively

Figure 5. Comparison of junction positions between the inverted repeats and single copy regions in a number of sequenced monocots. Included are *Najas major* along with three other alismatids: *Elodea canadensis* (JQ310743), *Sagittaria lichuanensis* (NC029815), *Potamogeton perfoliatus* (NC029814) and other related monocots: *Tofieldia thibetica* (NC_029813), *Lemna minor* (DQ400350), *Wolffiella lingulata* (NC015894), *Colocasia esculenta* (NC016753) and *Acorus calamus* (NC_007407).

Figure 6. Comparison of inverted repeat and single copy border positions across nine *Najas* plastid genomes (broken lines). Sizes in base pairs of each of the four major plastome components (LSC, IR_B, SSC, IR_A) are indicated. Lengths of intergenic regions adjacent to borders are given, along with the length of genes that span the borders. Pseudogenes are denoted by ψ.

Figure 7. Forward and palindromic repeat size and frequency in nine *Najas* taxa and related monocots identified with REPuter (Kurtz et al. 2001), at a repeat length of ≥30 bp with a Hamming distance of 3. Vertical bars represent repeats clustered in classes of 30-39, 40-49, 50-69 and 70-90. Scale bars differ between repeat types.

Figure 8. Inversion in large single copy region in *Hydrilla*. Exon two of trnV-UAC is missing in *Hydrilla* due to this inversion. The ndhC-ndhK-ndhJ operon is in the centre of this region. *Hydrilla* contains partial copies of ndhK and ndhJ genes, while *Najas* only contains a fragment of the ndhC gene. This region is shown in *Lemna minor*, *Elodea canadensis*, *Nicotiana tabacum* and *Najas major* for comparison.

Figure 9. Inversion in large single copy region in *Hydrilla* between accD and the psbE-psbJ operon. This region is shown in *Lemna minor*, *Elodea canadensis*, *Nicotiana tabacum* and *Najas major* for comparison.

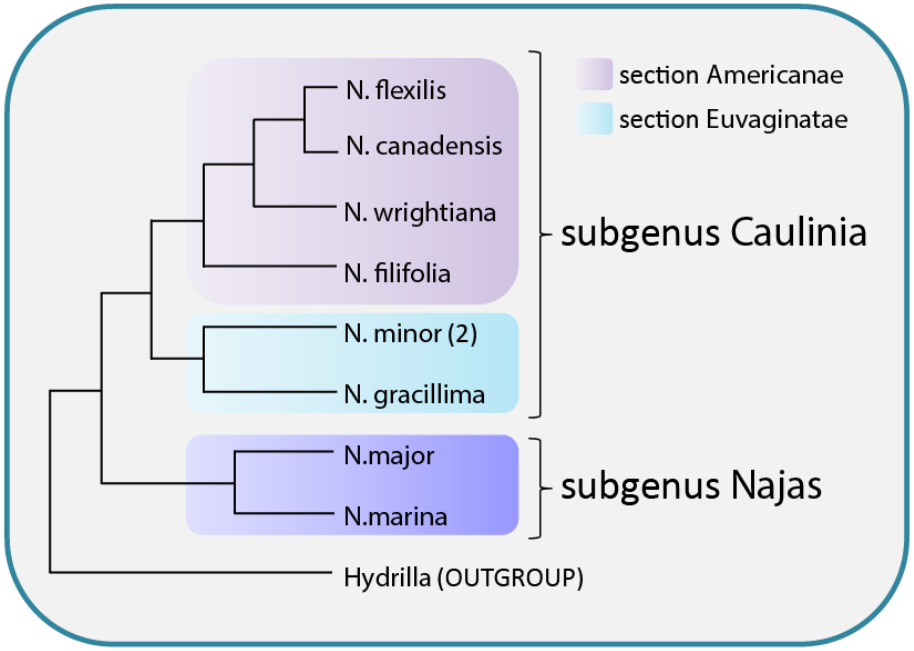

section Americanae
section Euvaginatae
N. flexilis
N. canadensis
N. wrightiana
N. filifolia
N. minor (2)
N. gracillima
N.major
N.marina
Hydrilla (OUTGROUP)
subgenus Caulinia
subgenus Najas

Figure 1

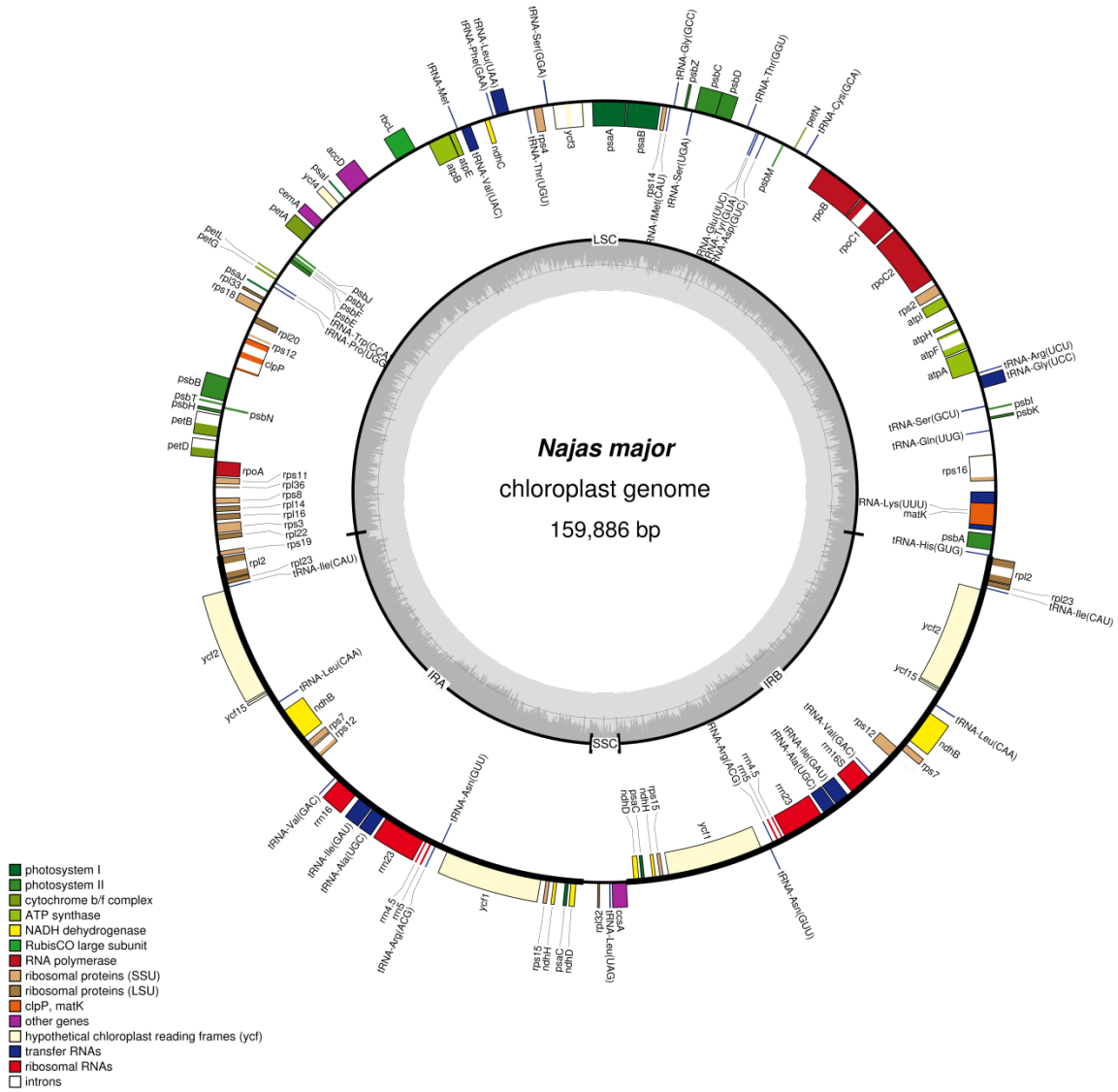
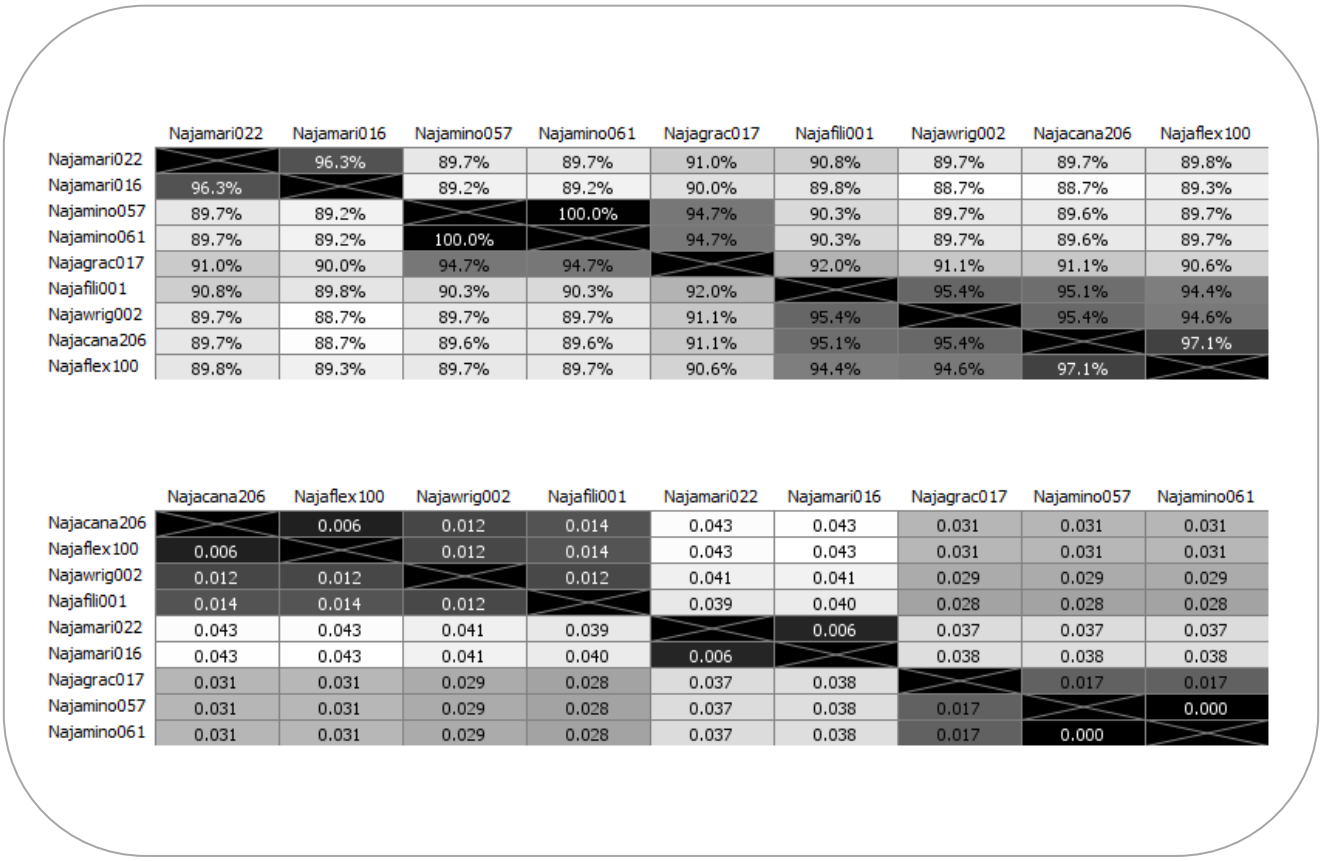

Najas major
chloroplast genome
159,886 bp
LSC
IRA
IRB
SSC
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs
introns

Figure 2



	Najamari022	Najamari016	Najamino057	Najamino061	Najagrac017	Najafili001	Najawrig002	Najacana206	Najaflex100
Najamari022		96.3%	89.7%	89.7%	91.0%	90.8%	89.7%	89.7%	89.8%
Najamari016	96.3%		89.2%	89.2%	90.0%	89.8%	88.7%	88.7%	89.3%
Najamino057	89.7%	89.2%		100.0%	94.7%	90.3%	89.7%	89.6%	89.7%
Najamino061	89.7%	89.2%	100.0%		94.7%	90.3%	89.7%	89.6%	89.7%
Najagrac017	91.0%	90.0%	94.7%	94.7%		92.0%	91.1%	91.1%	90.6%
Najafili001	90.8%	89.8%	90.3%	90.3%	92.0%		95.4%	95.1%	94.4%
Najawrig002	89.7%	88.7%	89.7%	89.7%	91.1%	95.4%		95.4%	94.6%
Najacana206	89.7%	88.7%	89.6%	89.6%	91.1%	95.1%	95.4%		97.1%
Najaflex100	89.8%	89.3%	89.7%	89.7%	90.6%	94.4%	94.6%	97.1%	

	Najacana206	Najaflex100	Najawrig002	Najafili001	Najamari022	Najamari016	Najagrac017	Najamino057	Najamino061
Najacana206		0.006	0.012	0.014	0.043	0.043	0.031	0.031	0.031
Najaflex100	0.006		0.012	0.014	0.043	0.043	0.031	0.031	0.031
Najawrig002	0.012	0.012		0.012	0.041	0.041	0.029	0.029	0.029
Najafili001	0.014	0.014	0.012		0.039	0.040	0.028	0.028	0.028
Najamari022	0.043	0.043	0.041	0.039		0.006	0.037	0.037	0.037
Najamari016	0.043	0.043	0.041	0.040	0.006		0.038	0.038	0.038
Najagrac017	0.031	0.031	0.029	0.028	0.037	0.038		0.017	0.017
Najamino057	0.031	0.031	0.029	0.028	0.037	0.038	0.017		0.000
Najamino061	0.031	0.031	0.029	0.028	0.037	0.038	0.017	0.000	

Figure 3

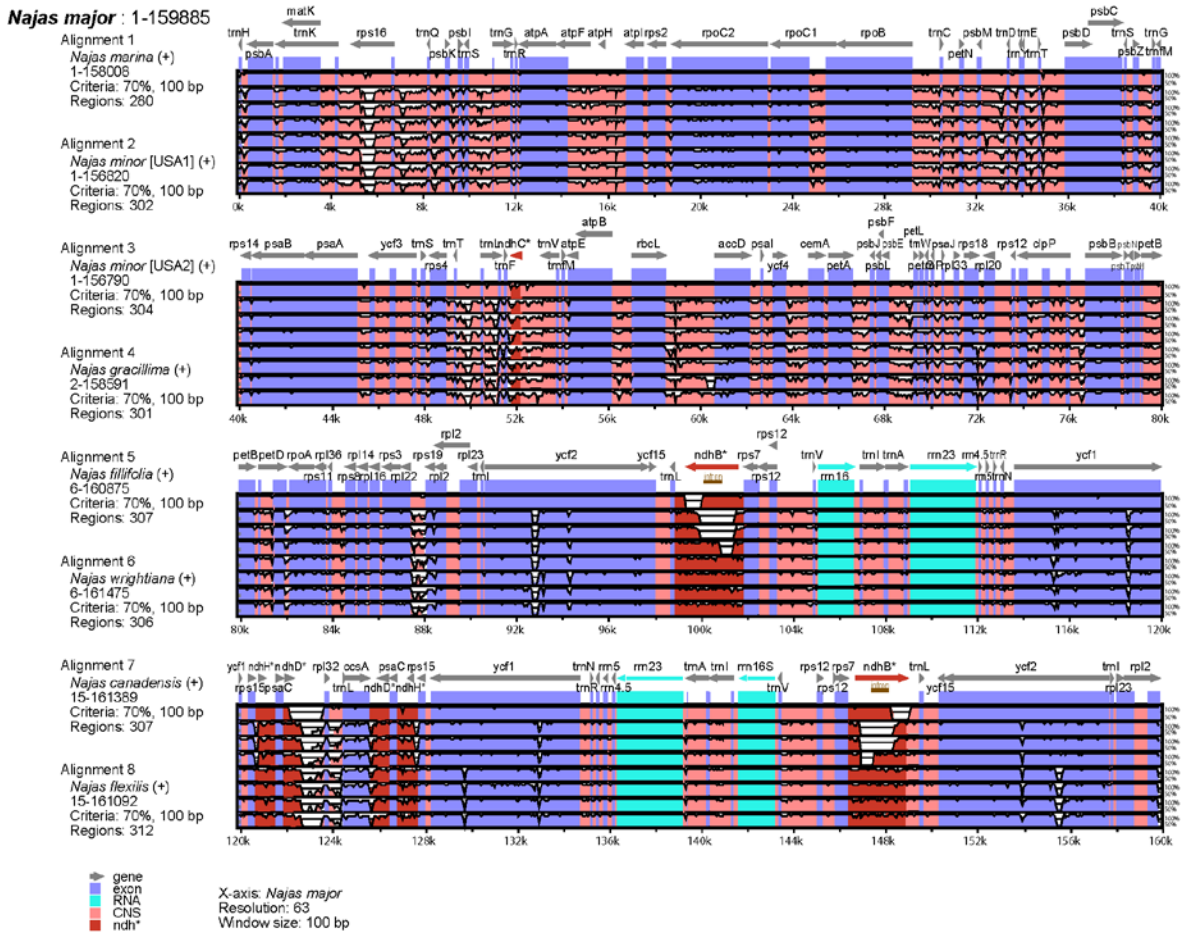

Najas major : 1-159885
Alignment 1
Najas marina (+)
1-158006
Criteria: 70%, 100 bp
Regions: 280
Alignment 2
Najas minor [USA1] (+)
1-156820
Criteria: 70%, 100 bp
Regions: 302
Alignment 3
Najas minor [USA2] (+)
1-156790
Criteria: 70%, 100 bp
Regions: 304
Alignment 4
Najas gracillima (+)
2-158591
Criteria: 70%, 100 bp
Regions: 301
Alignment 5
Najas fillifolia (+)
6-160875
Criteria: 70%, 100 bp
Regions: 307
Alignment 6
Najas wrightiana (+)
6-161475
Criteria: 70%, 100 bp
Regions: 306
Alignment 7
Najas canadensis (+)
15-161389
Criteria: 70%, 100 bp
Regions: 307
Alignment 8
Najas flexilis (+)
15-161092
Criteria: 70%, 100 bp
Regions: 312
gene
exon
RNA
CNS
ndh*
X-axis: Najas major
Resolution: 63
Window size: 100 bp

Figure 4

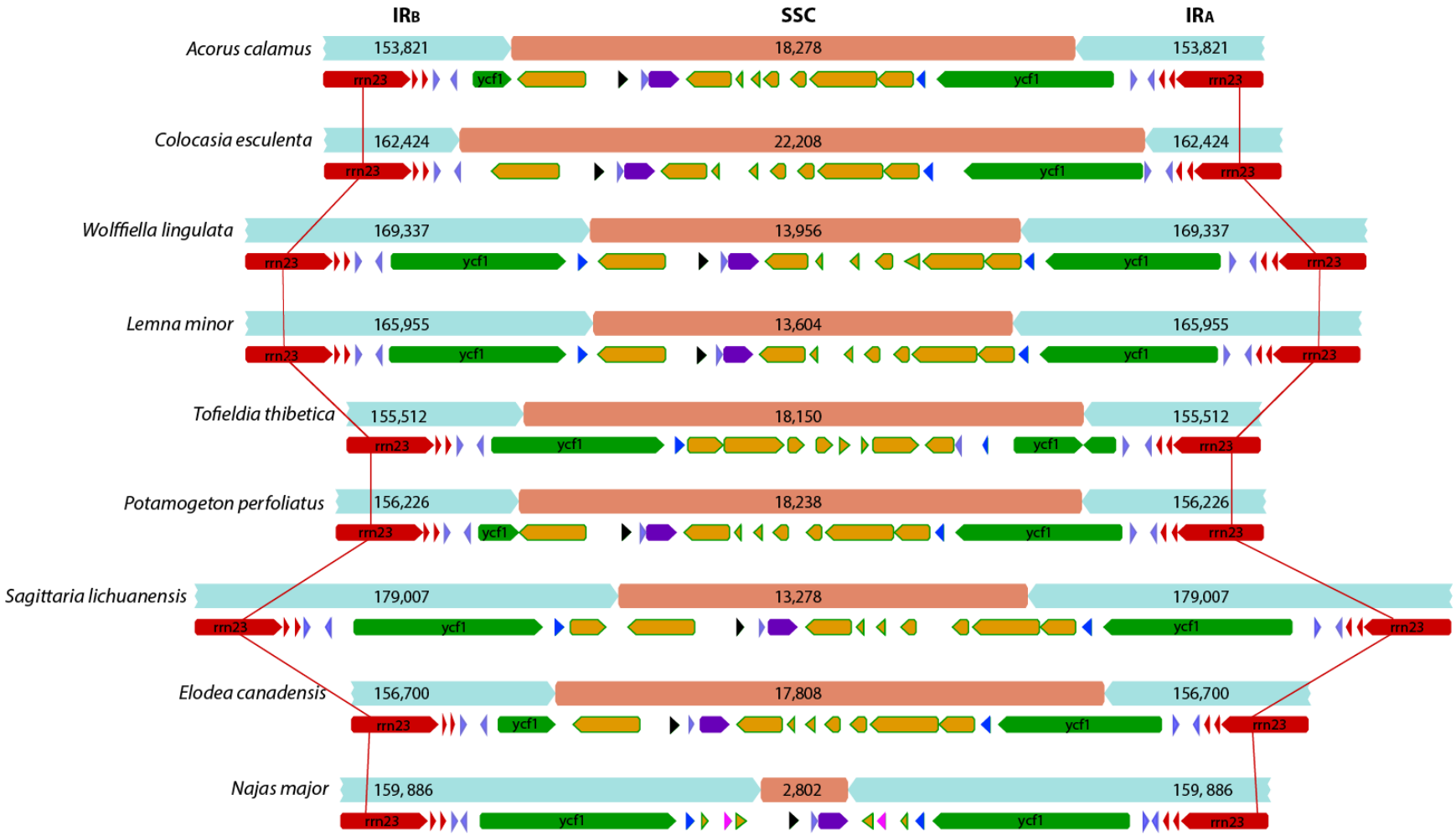

IRB
SSC
IRA
Acorus calamus
153,821
18,278
153,821
rrn23
ycf1
ycf1
rrn23
Colocasia esculenta
162,424
22,208
162,424
rrn23
ycf1
rrn23
Wolffiella lingulata
169,337
13,956
169,337
rrn23
ycf1
ycf1
rrn23
Lemna minor
165,955
13,604
165,955
rrn23
ycf1
ycf1
rrn23
Tofieldia thibetica
155,512
18,150
155,512
rrn23
ycf1
ycf1
rrn23
Potamogeton perfoliatus
156,226
18,238
156,226
rrn23
ycf1
ycf1
rrn23
Sagittaria lichuanensis
179,007
13,278
179,007
rrn23
ycf1
ycf1
rrn23
Elodea canadensis
156,700
17,808
156,700
rrn23
ycf1
ycf1
rrn23
Najas major
159, 886
2,802
159, 886
rrn23
ycf1
ycf1
rrn23

Figure 5

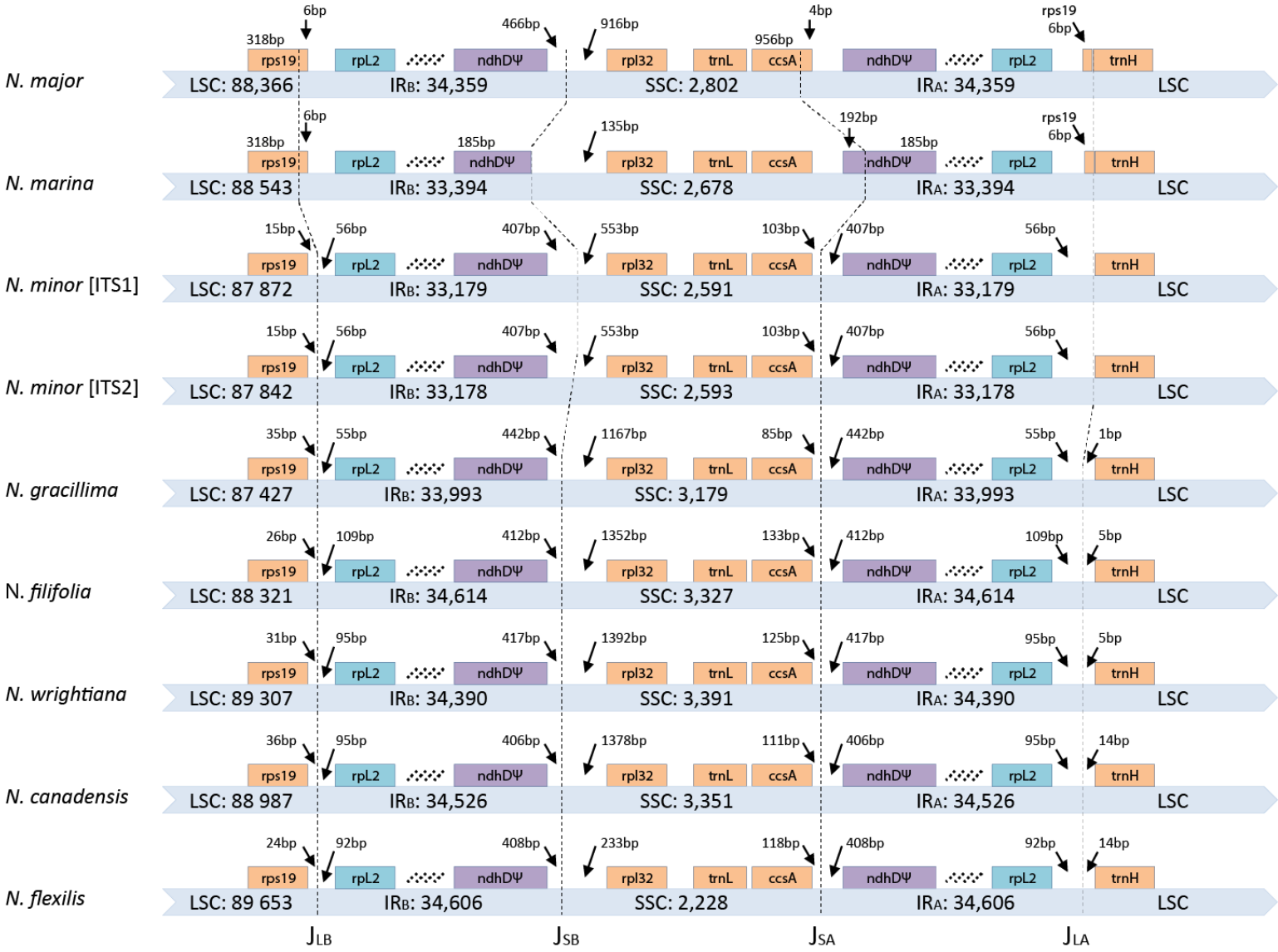

N. major
LSC: 88,366
IRB: 34,359
SSC: 2,802
IRA: 34,359
LSC
N. marina
LSC: 88 543
IRB: 33,394
SSC: 2,678
IRA: 33,394
LSC
N. minor [ITS1]
LSC: 87 872
IRB: 33,179
SSC: 2,591
IRA: 33,179
LSC
N. minor [ITS2]
LSC: 87 842
IRB: 33,178
SSC: 2,593
IRA: 33,178
LSC
N. gracillima
LSC: 87 427
IRB: 33,993
SSC: 3,179
IRA: 33,993
LSC
N. filifolia
LSC: 88 321
IRB: 34,614
SSC: 3,327
IRA: 34,614
LSC
N. wrightiana
LSC: 89 307
IRB: 34,390
SSC: 3,391
IRA: 34,390
LSC
N. canadensis
LSC: 88 987
IRB: 34,526
SSC: 3,351
IRA: 34,526
LSC
N. flexilis
LSC: 89 653
IRB: 34,606
SSC: 2,228
IRA: 34,606
LSC
rps19
rpL2
ndhDΨ
rpl32
trnL
ccsA
trnH
JLB
JSB
JSA
JLA

Figure 6

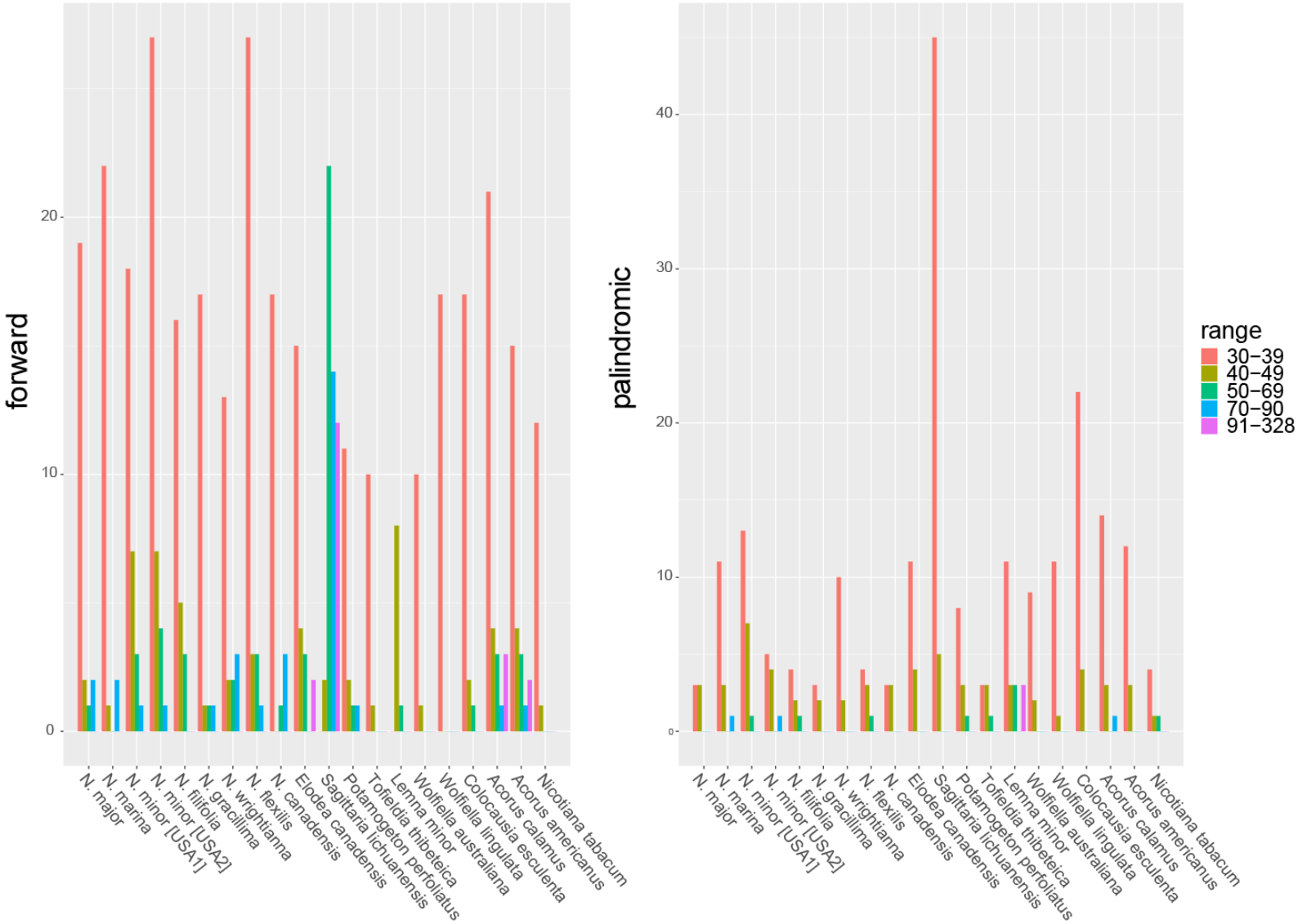

forward
palindromic
range
30–39
40–49
50–69
70–90
91–328
0
10
20
30
40
N. major
N. marina
N. minor [USA1]
N. minor [USA2]
N. filifolia
N. gracillima
N. wrightianna
N. flexilis
N. canadensis
Elodea canadensis
Sagittaria lichuanensis
Potamogeton perfoliatus
Tofieldia thibeteica
Lemna minor
Wolffiella australiana
Colocausia esculenta
Acorus calamus
Acorus americanus
Nicotiana tabacum

Figure 7

Hydrilla verticillata and *Najas major*
compared with *Lemna minor* (DQ400350), *Elodea canadensis* (NC_018541), and *Nicotiana tabacum* (Z00044)
rps4 to atpE region in LSC
(arrows indicate inverted region in *Hydrilla*)

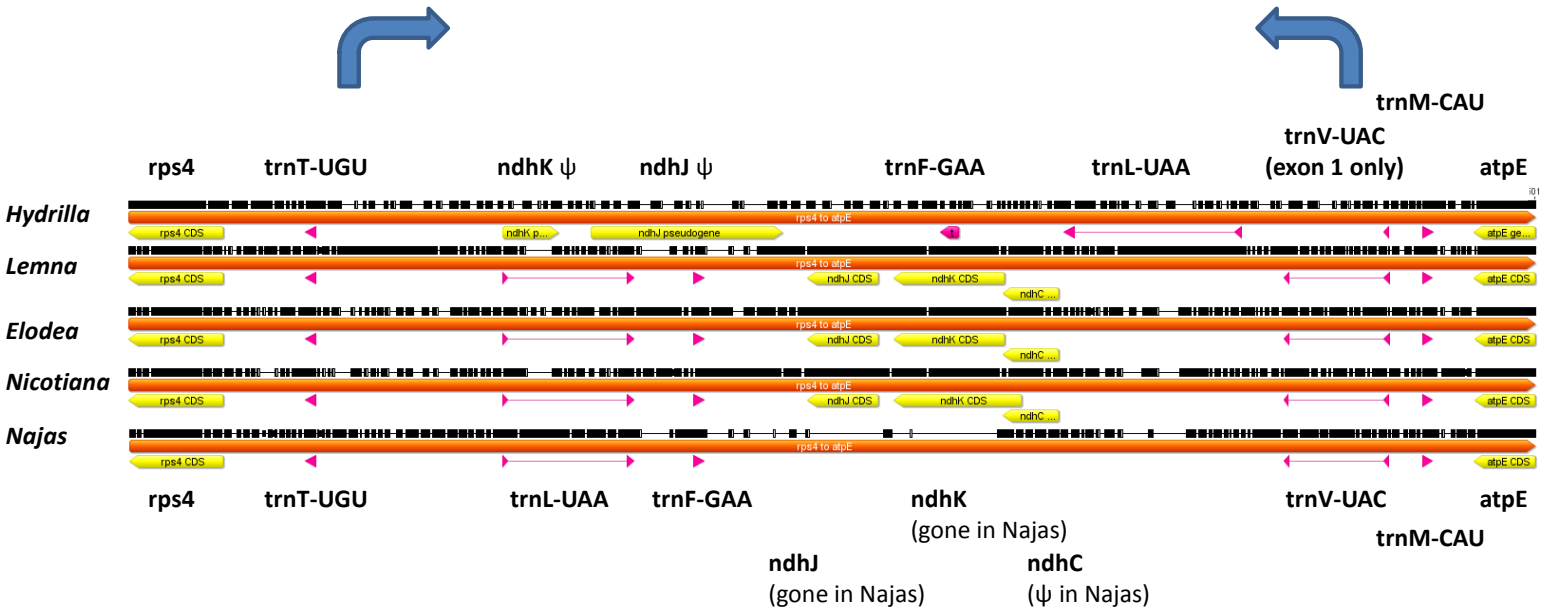


Figure 8

Hydrilla verticillata and *Najas major*
compared with *Lemna minor* (DQ400350), *Elodea canadensis* (NC_018541), and *Nicotiana tabacum* (Z00044)
accD to psbE region in LSC
(arrows indicate inverted region in *Hydrilla*)

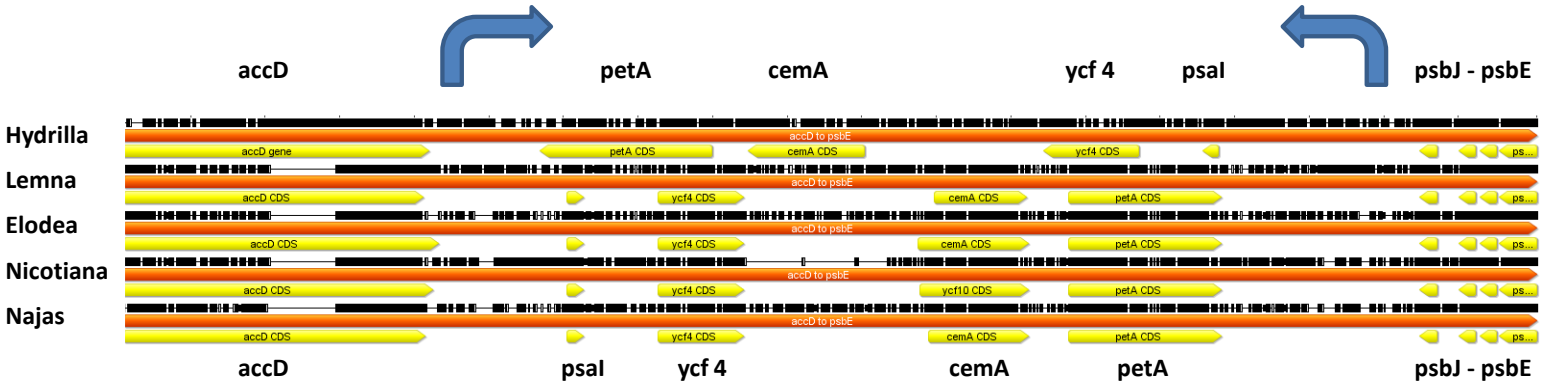


Figure 9

Chapter 4

Plastid phylogenomics of Hydrocharitaceae

Abstract

The frog's-bit, or tape-grass family (Hydrocharitaceae), is one of the largest families of aquatic flowering plants, containing both freshwater and marine species. In this family, we see the convergent evolution of unique pollination systems in some species, along with full adaptation to life in water in other species. Key to understanding the evolution of different traits within this group is the provision of a reliable phylogeny. In this study, which includes all genera in Hydrocharitaceae, with the exception of the rare Madagascan endemic, *Appertiella*, a plastid phylogenomic datsaset, which includes 63 protein coding genes and four ribosomal RNAs, is analyzed using maximum likelihood and Bayesian inference. Given that all genera within subfamily *Hydrilloideae* represent long branches in previous molecular analyses, I was interested in seeing the effect of different functional genes on relationships within the family. Therefore, along with analyzing all genes together in a concatenated dataset, four separate analyses were conducted based on different functional categories. In all of these analyses both *Hydrilla* and *Najas* resolved within subfamily *Hydrilloideae*; however, relationships within the family changed depending on the different functional class utilized. Additionally, even with the exclusion of genes which were interpreted as potentially containing conflicting phylogenetic signal, long branches to all taxa in *Hydrilloideae* were evident with both the full dataset and in trees constructed from the individual functional classes.

Introduction

Hydrocharitaceae

A high degree of morphological and anatomical diversity, overlaid with parallel convergences in vegetative morphology, anatomy and pollination systems have hampered previous attempts to recognize distinct patterns of relationships within the aquatic family Hydrocharitaceae (Juss.), based on these criteria (Tanaka et al. 1997, Les et al. 2006). With four subfamilies now recognized (reviewed in Les et al. 2006) (Figure 1 [A]), molecular phylogenies to date have been hampered by long branches, short internal nodes, and poor support characterizing some of the groupings within, and relationships between, these subfamilies (Iles et al. 2013). Two taxa within subfamily *Hydrilloideae* that consistently represent long branches in molecular studies, based on chloroplast regions, are the cosmopolitan genera, *Hydrilla* and *Najas* (Tanaka et al. 1997, Iles et al. 2013, Lou et al. 2015, Ross et al. 2016).

Traditionally regarded as a monotypic genus (comprising the single species, *Hydrilla verticillata*), recent molecular analyses have revealed a number of cryptic lineages within the predominantly clonal *Hydrilla* (Benoit 2011, Zhu et al. 2015). On the other hand, the annual genus *Najas* L., consists of approximately 40 species (Triest 1988), separated into two subgenera, *Najas* and *Caulinia* (Ascherson 1864, Rendle 1899). Although *Hydrilla* has had a long association with other 'hydrocharits', *Najas* historically typified the order Najadales Dumort. along with other aquatic taxa that are now placed in Potamogetonaceae, Ruppiaceae and

Zosteraceae. The preponderance of morphological (Shaffer-Fehre, 1991a, 1991b, Les et al. 2006) and molecular data (Les et al. 1993, 1997, Tanaka et al. 1997, Les et al. 2006, Petersen et al. 2006, Iles et al. 2013, Les & Tippery 2013) led Les & Tippery (2013) to dissolve the Najadales and place *Najas* within the alismatid family Hydrocharitaceae. With the exception of Li & Zhou (2009) (for discussion of homoplasious characters used in that study see Les & Tippery, 2013), all recent molecular data have continued to support that decision (Iles et al. 2013, Ross et al. 2016, Petersen et al. 2016).

Within *Najas*, subgenus *Najas* (containing *Najas major and N. marina*) is distinguished from subgenus *Caulinia* in containing dioecious brackish water species, having spines on the abaxial surface of the lamina and on the internodes, and having a seed coat four or more layers thick. All species within *Caulinia* are freshwater, principally monoecious, lacking spines on the internodes or undersides of the leaves, and having a seed testa which is three cell layers thick (Rendle 1899, Triest 1988). Apart from Chen et al. (2012), molecular studies to date have provided strong support for the monophyly of both subgenera (Les et al. 2010, Ito et al. 2017).

Reduced sequencing costs are increasingly prompting researchers to generate phylogenomic datasets from plastid coding regions, in an attempt to resolve relationships at both deep (e.g., Moore et al. 2010, Drew et al. 2014, Yan et al. 2015, Sun et al. 2016) and shallow scales (e.g., Wysocki et al. 2015, Harrison et al. 2016). The main strength of this approach is the elimination of random error in phylogenetic inference, associated with sampling just a few chloroplast loci (Philippe et al. 2011). Increasingly, however, it is recognized that adding more data, in itself, is not sufficient to resolve incongruence, and there may be many pitfalls associated with the simple addition of additional sequences (reviewed in Philippe et al. 2011). Conceivably, misleading signal may be present in large datasets, which may impact phylogenetic resolution,

and may be particularly evident in lineages that have experienced long genetic isolation, rapid radiation or divergence, or elevated evolutionary rates (Parks et al. 2012). Additionally, it is suggested that the use of genome-scale data, where taxon sampling is poor, but character sampling is rich, may be particularly susceptible to long-branch attraction (Leebens-Mack et al. 2005).

Recently, Ross et al. (2016) generated a plastid phylogenomic dataset for the Alismatales, which included all hydrocharit genera with the exception of *Appertiella* (endemic to Madagascar and previously unsampled in any molecular analyses) and *Hydrilla*. In this chapter, the plastid coding genes from *Hydrilla verticillata* and nine *Najas* taxa (Figure 1 [B]) are combined with those from the other hydrocharit taxa (Ross et al. 2016) in a phylogenetic analysis.

Some questions remaining to be resolved are 1) whether inclusion of *Hydrilla* will affect the placement of *Najas* (or other taxa) within Hydrocharitaceae, 2) whether a whole plastid gene set can provide enough phylogenetic signal to resolve relationships within *Najas* and 3) whether analyses based on different functional chloroplast genes will produce consistent relationships within Hydrocharitaceae.

Materials and Methods

Sampling

All protein-coding and ribosomal RNA (rRNA) genes for Hydrocharitaceae and Butomaceae were extracted from the plastid alignment of Ross et al. (2016). Individual gene regions were aligned separately, with corresponding regions included from *Hydrilla verticillata*, and the nine newly sequenced *Najas* plastomes (Chapter 3). As Ross et al. (2016) included a sample of *N. guadalupensis* in their analysis, this alignment represents all 16 hydrocharit genera represented

in Figure 1 [A], and 10 *Najas* taxa in total. *Butomus umbellatus* was used as the outgroup for Hydrocharitaceae, following Les & Tippery (2013).

Alignments

Protein-coding genes were aligned in MAFFT, with the translation align function (scoring matrix: Blosum62, Gap open penalty: 1.53 and offset value 0.123), and the four rRNA genes were aligned with the FFT-NS-i x1000 algorithm, (scoring matrix 100PAM/k=2, gap open penalty 2.01, with an offset value of 0.123). Alignments were visually inspected and manually adjusted where necessary in Geneious V6.1 (http://www.geneious.com, Kearse et al. 2012).

As attempts to confidently align certain gene regions even within *Najas* proved difficult, I chose to employ a conservative strategy (outlined in Appendix A). Briefly, the 3' end of eight genes was removed manually, 15 genes were trimmed with Gblocks v0.91 (Talavera & Castresana 2007), implemented in TranslatorX (Abascal et al. 2010), using the bacterial genetic code with less stringent settings. Finally, 16 genes were removed entirely from the analysis. All 11 *ndh* genes were excluded, as these genes are either lost or pseudogenized in all members of subfamily *Hydrilloideae* (Iles et al. 2013, Peredo et al. 2013, Ross et al. 2016 and Chapter 3 here), along with a further five genes (*accD*, *clpP*, *infA*, *ycf1* and *ycf2*). Additionally, the matK gene for *Nechamandra* and *Vallisneria* was removed and coded as missing data for these taxa. This resulted in a total of 63 protein-coding and 4 rRNA gene alignments, which were concatenated for the full Hydrocharitaceae dataset.

Apart from determining relationships based on this full Hydrocharit dataset, to further explore phylogenetic signal with respect to relationships within *Najas*, a number of different analyses were undertaken. As *Najas* represents a long branch in previous phylogenetic analyses,

and either resolves as sister to the whole *Hydrilloideae* clade, sister to *Hydrilla*, *Nechamandra* and *Vallisneria*, or as sister to *Hydrilla* alone (reviewed in Les & Tippery 2013), I was interested in the effect of different outgroups on relationships within *Najas* (e.g., Heath et al. 2008). Consequently, to evaluate the effect of taxon sampling on relationships within *Najas*, a series of analyses were run in which different hydrocharit taxa were used as outgroups for *Najas*.

As I was interested in the effects of different functional chloroplast genes on phylogenetic relationships in the family (and in particular whether branch topology might change with the analysis of different classes), four separate analyses (including all taxa) were run, using the photosynthesis, transcription, ribosomal protein and rRNA genes (see Appendix B for list of genes).

Phylogenetic analysis

For all analyses, partition schemes and suitable models were initially evaluated in PartitionFinder v1.1.1 (Lanfear et al. 2014), using the Bayesian information criterion (BIC) (Schwarz 1978, Sullivan & Joyce 2005). The following *a priori* partitioning schemes were assessed a) single partition, b) by gene, with a separate partition for rRNA, c) by codon position, with a separate partition for rRNA and d) by gene and codon position, with a separate partition for rRNA. Phylogenetic trees were inferred using maximum likelihood (ML) in IQ-TREE (V. 1.5.beta 4) (Nguyen et al. 2015), using the suggested partitions and models from PartitionFinder, with each partition allowed to have its own evolutionary rate (–spp option) (Chernomor et al. 2016). Branch support was assessed using the ultrafast bootstrap method (UFBoot2) of Hoang et al. (2017), and support was then compared with results from standard nonparametric bootstrapping (1000 replicates), also implemented in IQ-TREE. Additionally, a number of analyses were

repeated to evaluate whether stochasticity in the method resulted in trees being stuck in local optima.

Bayesian MCMC analyses were carried out in MrBayes 3.2.4 (Ronquist & Huelsenbeck 2003), with two independent runs of either 20 or 30 million generations, each with four simultaneous chains, and a discard burn-in of 25%. Initial analyses suggested that 20 million generations was sufficient for good mixing and run convergence for all datasets, with the exception of the rRNA which were run for 80 million generations. Priors were left at their default setting, with the exception of the branch length prior which was a more diffuse compound Dirichlet prior [GammaDir(1.0,0.100,1.0,1.0)] (Rannala et al. 2012, Lewis et al. 2016). With this prior, tree length is associated with a gamma distribution with a mean of 10 and branch length proportions are associated with a uniform Dirichlet distribution. A variable rate multiplier prior was applied across partitions (Dirichlet 1.00, 1.00), and all parameters were unlinked across partitions with the exception of branch lengths and topology (see Appendix C for parameters). Effective sample sizes (>200), mixing and convergence of parameter values were assessed in Tracer v1.5 (Rambaut and Drummond 2007). Bayesian analyses were run on the University of Connecticut's BBC Bioinformatics Facility Cluster.

Results

Full Hydrocharitaceae dataset

The final aligned dataset, including all 26 taxa, was 45816 bp comprising 63 protein-coding genes and the four rRNAs. Optimal partitioning in PartitionFinder, based on the BIC score, for the full gene dataset reflected codon position, with the four rRNAs as a separate partition (Appendix C).

In the ML analyses, the same tree topologies, and similar tree likelihoods, were obtained with separate independent runs of IQ-TREE. The rapid bootstrapping method (UFboot2) gave similar support values to the standard bootstrapping method, based on 1000 bootstraps, but with slightly elevated values on some nodes (Figure 2). However, application of this method comes with the proviso that one should only consider branches with greater than 90% support (Hoang et al. 2017).

In the full gene dataset, with all taxa included, the same topologies were obtained with both ML and Bayesian inference (BI), and support values were similar (Figure 2). Even with the removal of genes that were suspected to have unusual substitution rates or conflicting phylogenetic signal (approximately 40% of the original Hydrocharitaceae dataset), long branches to all genera in *Hydrilloideae* were still evident in these analyses.

Hydrilla resolves as sister to *Nechamandra* and *Vallisneria*, but support for the monophyly of the seagrass/*Nechamandra*/*Vallisneria* clade, and the *Enhalus*/ *Thalassia* clade, is slightly diminished in the ML analysis from that of Ross et al. (2016). Additionally, the position of *Stratiotes* still remains under question with these data; as here, *Stratiotes* resolves as sister to the *Anacharidoideae*, as opposed to sister to the rest of Hydrocharitaceae (Ross et al. 2016).

Analysis of relationships within *Najas*

Support at the base of the *Najas* clade is poor, with this large chloroplast dataset unable to completely resolve relationships within the genus (Figure 2). Subgenus *Najas* (*N. major* and *N. marina*) resolves as a distinct clade, but support for subgenus *Caulinia* as a cohesive unit is lacking. Relationships within section *Americanae* and section *Euvaginatae* are well-supported and what we would expect from previous data (Les et al. 2010).

To investigate whether relationships within *Najas* would change if different hydrocharit outgroups were used, a number of analyses were run including different taxa as outgroups. No substantial increase in support for the monophyly of subgenus *Caulinia* was observed in these analyses (Figure 3). Initial analyses of just the *Najas* taxa, without inclusion of an outgroup, resulted in a total of 1230 parsimony informative characters (Appendix C), and midpoint rooted trees resulted in full support in both ML and BI for subgenus *Caulinia*. Additionally, although with negligible support, subgenus *Najas* was observed to move position within the genus, depending on the incorporated outgroup; and this was not consistent between the phylogenetic methods. Separate analyses with the exclusion of these taxon groups (data not shown) resulted in a similar outcome. Given the morphological and ecological distinctness of subgenus *Najas*, the relationship we expect is one where subgenus *Najas* is sister to subgenus *Caulinia*, wherein section *Americanae* and section *Euvaginate* are a well-supported clade (as in the analysis with only the *Najas* taxa); however, none of these analyses could support this hypothesis.

Chloroplast functional class genes

Alignments of the genes for the four functional classes were comprised of 22878 bp (photosynthesis genes), 9264 bp (PEP polymerase genes), 7551 bp (ribosomal proteins) and 6123 bp (rRNA genes). Evaluation of partitions and models in PartitionFinder for the four functional gene class alignments suggested small changes to the partitioning schemes and models for some of the gene classes, from those employed on the full gene dataset. For example, rather than partitioning by each codon position, first and second codon positions were grouped into a single subset for the PEP polymerase genes; and the suggested model for the single partition of the rRNA genes was GTR with invariant sites, rather than the GTR plus gamma model (see Appendix C for model parameters and parsimony informative characters for each gene class).

The photosynthesis genes were the only group that strictly reflected the partitions and models suggested for the full dataset.

The same topologies resulted from both ML and BI analyses, with similar support values obtained with these two methods for each of the four functional classes of genes (Figure 4). In these analyses however, topologies and relative branch lengths were observed to change across the four functional groups. Relationships in the family were similar with the photosynthesis, PEP polymerase and ribosomal proteins, with the exception of the placement of *Stratiotes*, which variably resolved as sister to the *Anacharidoideae* or as sister to the rest of the family, either with good support (photosynthesis genes) or with poor support (PEP polymerase and ribosomal proteins). Of these three gene groups, branch lengths to *Hydrilloideae* produced with both the PEP polymerase and ribosomal protein genes were relatively longer than those with the photosynthesis genes.

The greatest overall differences were observed in relationships produced with the four rRNA genes, with a number of topological incongruences between trees produced with these genes and those produced by the other three gene groups (and the full gene dataset). Additionally, much longer branches leading to taxa within subfamily *Hydrilloideae*, relative to those leading to taxa in the other three subfamilies, were observed. Within the *Anacharidoideae*, short branches were evident, as expected, with these highly conserved genes; and *Lagarosiphon* still resolved within this group (with good support). However, with respect to the other taxa, a number of relationships changed (albeit with varying support); for example, *Limnobium* and *Hydrocharis* resolved with the *Anacharidoideae* (ML:54, BI:0.99), *Najas* resolved as sister to *Hydrilla*, *Nechamandra* and *Vallisneria* (ML: 88, BI: 1.00), and *Halophila* (which represents a very long branch with these data) resolved in a clade with *Enhalus*, with full support.

It was observed that Bayesian analyses of the rRNA genes experienced difficulty in reaching stationarity, and good effective sample sizes. Additionally, conflicting tree lengths for these genes were produced by ML and BI (Appendix C), with the Bayesian analysis resulting in a longer tree solution. Similar topologies were sampled by both paired runs in these analyses, reflected by the average standard deviation of split frequency values (0.001); and the potential scale reduction factor values (1.00) indicated convergence of final mean parameter estimates. However, even at 60 million generations, effective sample sizes for the rate multiplier and tree length were low, and did not reach a sample size of 200 until over 70 million generations. Additionally, it was observed that acceptable rates (~20-60%) for branch swaps between the heated and cold chains were not achieved, indicating that the MCMC algorithm had not adequately sampled regions of high posterior probability. The suggested model for these genes was a GTR model with invariant sites; however, I also ran these data with a GTR plus gamma model, with similar results. The same branch length prior was used for all the Bayesian analyses (GammaDir[1.0,0.100,1.0,1.0]). These compound Dirichlet priors are estimated to be more robust and less prone to branch-length overestimation than default exponential priors in MrBayes (Rannala et al. 2012). This behavior in the Bayesian analysis was worrisome and will require further investigation. Further analysis of this data may benefit from heating the Metropolis-coupled chains to encourage movements between isolated peaks in the posterior distribution, or increasing the number of incrementally heated chains, or frequency of chain swaps.

Discussion

Relationships within Hydrocharitaceae

Earlier morphological analyses placed *Hydrilla* in a clade with the anacharit genera, principally due to convergent vegetative morphologies (reviewed in Les et al. 2006). Both the full chloroplast gene dataset and the individual chloroplast functional groups analyzed here support the inclusion of *Hydrilla* in subfamily *Hydrilloideae*, consistent with previous molecular analyses (reviewed in Les & Tippery 2013 and Iles et al. 2013). In all analyses, *Hydrilla* resolved as sister to *Nechamandra* and *Vallisneria*; however, the inclusion of *Hydrilla* results in support for the grouping of *Nechamandra*/*Vallisneria* and the seagrasses being slightly diminished in the maximum likelihood analysis from that obtained by Ross et al. (2016). Additionally, the picture becomes a little more complicated for this group when the chloroplast genes are analyzed as separate functional classes. Although molecular analyses consistently have placed *Najas* within the subfamily *Hydrilloideae*, the exact position within this subfamily is still questioned (Les & Tippery 2013). With these data, the position of *Najas* within this subfamily continues to be supported with the inclusion of *Hydrilla*; however, a conflicting placement of *Najas* either as sister to the rest of the subfamily (photosynthesis, PEP polymerase and ribosomal protein genes), or as sister to the *Hydrilla*/*Nechamandra*/*Vallisneria* clade (rRNA genes), results from analyses of the different functional groups.

The position of *Stratiotes* within Hydrocharitaceae remains intractable with these data, with insufficient characters available in this plastome dataset to resolve the relationship of this taxon to the rest of the groups with confidence.

Relationships within *Najas*

Clearly, relationships within *Najas* have not been resolved either with these data. As our focus is on North American *Najas*, we have a good representation of diversity within section *Americanae*

included here. Additionally, only two species are recognized in subgenus *Najas* (Rüegg et al. 2016), and we have included both of these in our analysis. Incorporation of other samples from section *Euvaginatae*, *Spathaceae* and *Nudae* may help to improve resolution within *Najas*; however, the long branches to *Najas* and the other genera in *Hydrilloideae*, may still leave a problem with regards choosing a suitable outgroup. The other genera in *Hydrilloideae* are species poor (Cook 1982, Les 1988), and with the exception of *Vallisneria*, little additional taxon sampling will be available that might allow breaking up these long branches. Inclusion of more variable chloroplast intergenic and intron regions would provide more markers at a shallow scale, but the problem of aligning these regions across taxa at this divergence then becomes much more difficult.

Pruning the original dataset of ambiguously aligned regions, may have removed many of the more informative sites within *Najas* (see Tan et al. 2015 for discussion on alignment filtering). However, in handling this dataset I attempted to be as conservative as possible, choosing to manually remove just the variable 3' end of genes (which I had observed to be highly variable across *Najas* in initial alignments of just that group), rather than submitting entire genes to GBlocks which would have removed more of the variable regions. This method is, of course, also questionable in its subjectivity and lack of repeatability.

Recently, Ito et al. (2017) chose to collapse sections *Euvaginatae*, *Spathaceae* and *Nudae*, based on a molecular analysis of four plastid genes (rbcL, matK, rpoA and rpoC1) and nrITS. An unusual result in their analysis was the grouping of *N. chinensis* (along with two other unidentified species) within section *Americanae* with strong support, along with the polyphyletic positioning of samples of *N. graminea* and *N. tenuifolia*. Given the results here, perhaps caution

is required in reversing previous taxonomic decisions at this level, until a greater sampling can be achieved.

Chloroplast functional class genes

Although, for the most part, representing a single linkage group, different regions of the plastome are not exempt from exhibiting mutational rate variation which may provide misleading phylogenetic signal (Goremykin et al. 2010), and disproportionally influence the topology of weakly supported nodes (Parks et al. 2012). Some researchers have attempted to investigate this by using methods such as removing the most saturated sites (for example, Goremykin et al. (2010, 2013, 2015), Parks et al. (2012), Rajan (2013), but see Drew et al. 2014), or long branch taxa (Parks et al. 2012), or by increasing taxon sampling (Leebens-Mack et al. 2005, Heath et al. 2008).

As a single molecule, the chloroplast should represent one history (in the absence of biparental inheritance, heteroplasmy, or recombination), and we expect the phylogenetic signal to be similar across all the genes; however, concatenating genes into one dataset can often mask the effect of any opposing signal in different genes (Lewis et al. 2016). Analysis of the different functional gene classes in Hydrocharitaceae further demonstrates that caution is required in handling these datasets.

Najas plastome genes have already been incorporated in phylogenomic analyses to infer relationships at deeper levels within the Alismatales and monocots (e.g., Ross et al. 2016, Luo et al. 2016), with the results of these analyses having important implications for character evolution in monocots (Rudall et al. 2017). In Chapter 3, I discussed genes that have unusual evolutionary

patterns in both *Najas* and *Hydrilla* (and perhaps other *Hydrilloideae* taxa: Appendix C), and a number of these genes were excluded from our analysis here (e.g., accD, clpP1, infA). Additionally, Chapter 2 shows that at least some of the polymerase genes in *Najas* have elevated dn/ds ratios. If these genes are under positive selection in *Najas* (and other Hydrocharitaceae) then inclusion of these genes would violate assumptions of any substitution model. If, on the other hand, they are under relaxed selection, their behavior in phylogenetic analyses needs to be evaluated further.

References

Abascal, F. Zardoya, R. & M. J. Telford. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. 2: 7-13.

Ascherson, P. 1864. Flora der Provinz Brandenburg 1 (2). Hirschwald, Berlin. http://bibdigital.rjb.csic.es/ing/Libro.php?Libro=2953

Benoit, L.K. 2011. Cryptic speciation, genetic diversity and herbicide resistance in the invasive aquatic plant *Hydrilla verticillata* (L.f.) Royle (Hydrocharitaceae). Ph.D. dissertation, University of Connecticut, Storrs, CT. Website (http://digitalcommons.uconn.edu/dissertations/AAI3492149).

Chen, L.Y. Chen, J.M. Gituru, R.W. & Q.F. Wang. 2012. Generic phylogeny, historical biogeography and character evolution of the cosmopolitan aquatic plant family Hydrocharitaceae. BMC Evolutionary Biology. 12: 30.

Chernomor, O. von Haeseler, A. & B. Q. Minh. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol. 65:997–1008.

Cook, C.D.K. & R. Luond. 1982. A revision of the genus *Hydrilla* (Hydrocharitaceae). Aquatic Botany 13:485–504.

Davis, J.I. McNeal, J.R. Barrett, C.F. et al. 2013. Contrasting patterns of support among plastid genes and genomes for major clades of the monocotyledons. In:Wilkin P, Mayo SJ, eds. Early events in monocot evolution. Systematics Association Special Volume Series. Cambridge: Cambridge University Press, 315–349.

Drew B.T., Ruhfel B.R., Smith S.A., Moore M.J., Briggs B.G., Gitzendanner M.A., Soltis P.S., Soltis D.E. 2014. Another look at the root of the angiosperms reveals a familiar tale. Syst. Biol. 63:368–382.

Goremykin, V. Nikiforova, S. & O. Bininda-Emonds. 2010. Automated removal of noisy data in phylogenomic analyses. J Mol Evol. 71:319–331.

Goremykin V.V. Nikiforova S.V Biggs P.J. Zhong B. Delange P. Martin W. Woetzel S. Atherton R.A. McLenchan P.A. & P.J. Lockhart. 2013. The evolutionary root of flowering plants. Syst. Biol. 62:50–61

Goremykin, V.V. Nikiforova, S.V. Cavalieri, D. Pindo, M. & P. Lockhart. 2015. The Root of Flowering Plants and Total Evidence, Syst. Biol. 64: 879–891

Harrison, N. Harrison, R.J. & C.A. Kidner. 2016. Comparative Analysis of Begonia Plastid Genomes and Their Utility for Species-Level Phylogenetics. PloS ONE 11: e0153248. doi:10.1371/journal. pone.0153248

Heath, T.A. Hedtke, S.M. & M. Hillis. 2008. Taxon sampling and the accuracy of phylogenetic analyses, J. Syst. Evol. 46: 239–257.

Hoang, D.T. Chernomor, O. von Haeseler, A. Minh, B.Q. & L.S. Vinh. 2017. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. doi: 10.1093/molbev/msx281.

Iles, W.J. Smith, S.Y. & S.W. Graham. 2013. A well-supported phylogenetic framework for the monocot order Alismatales reveals multiple losses of the plastid NADH dehydrogenase complex and a strong long-branch effect. *In* Wilkin, P. Mayo, S.J. [eds.], Early Events in Monocot Evolution, 1–28. Cambridge University Press Systematics Association Special Series, Cambridge, UK.

Ito, Y. Tanaka, N. Gale, S. Yano, O. & Li, J. 2017. Phylogeny of *Najas* (Hydrocharitaceae) revisited: Implications for systematics and evolution. Taxon, 66: 309–323.

Kearse, M.R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649.

Lanfear, R.B. Calcott, D. Kainer, C. Mayer & A. Stamatakis. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC evolutionary biology 14: 82.

Leebens-Mack, J., Raubeson, L.A., Cui, L., Kuehl, J.V., Fourcade, M.H.,Chumley, T.W., Boore, J.L., Jansen, R.K., & C.W. dePamphilis. 2005. Identifying the basal angiosperm node in chloroplast genome phylogenies: Sampling one's way out of the Felsenstein zone. Mol.Biol. Evol.22:1948–1963.

Leigh, J. Susko, E. Baumgartner M. & A. Roger. 2008. Testing congruence in phylogenomic analysis. Syst. Biol. 57:104–115.

Les, D. H. 1988. Breeding systems, population structure, and evolution in hydrophilous angiosperms, 75:819–835.

Les, D.H. Garvin, D.K. and Wimpee, C.F. 1993. Phylogenetic studies in the monocot subclass Alismatidae: evidence for a reappraisal of the aquatic order Najadales. Molecular Phylogenetics and Evolution, 2, 304–314.

Les, D.H. M.L. Moody & C.L. Soros. 2006. A reappraisal of the phylogenetic relationship in the monocotyledon family Hydrocharitaceae (Alismatidae). Aliso 22: 211–230.

Les, D.H. S.P. Sheldon & N.P. Tippery. 2010. Hybridization in Hydrophiles: Natural Interspecific Hybrids in *Najas* (Hydrocharitaceae). Systematic Botany 35: 736–744.

Les, D.H. & N.P. Tippery. 2013. In time and with water... the systematics of alismatid monocotyledons. *In* Wilkin, P. Mayo, S.J. [eds.], Early Events in Monocot Evolution, 118–164. Cambridge University Press Systematics Association Special Series, Cambridge, UK.

Les, D.H. Peredo, E.L. Tippery, N.P. Benoit, L.K. Razifard, H. King, U.M. Na, H.R. et al. 2015. *Najas minor* (Hydrocharitaceae) in North America: A reappraisal. Aquatic Botany 126: 60–72.

Lewis, P. O. M.-H. Chen, L. Luo, L.A. Lewis, K. Fučiková, S. Neupane, Y.-B. Wang, D. Shi. 2016. Estimating bayesian phylogenetic information content Syst. Biol. 65: 1009–1023.

Li, X. & Z. Zhou. 2009. Phylogenetic studies of the core Alismatales inferred from morphology and rbcL sequences. Progress in Natural Science. 19: 931–945.

Luo, Y. Ma, P.F. Li, H.T. Yang, J.B. Wang, H. & D. Z. Li. 2016. Plastid phylogenomic analyses resolve Tofieldiaceae as the root of the early diverging monocot order Alismatales. Genome Biology and Evolution 8: 932–945.

Moore, M.J. Soltis, P.S. Bell, C.D. Burleigh, J.G. & Soltis, D.E. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc. Natl Acad. Sci. USA 107, 4623–4628.

Nguyen, L.T. Schmidt, H.A. vonn Haeseler A. & B.Q. Minh. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol Biol Evol, 32:268–274.

Parks, M. Cronn, R. & A. Liston. 2012. Separating the wheat from the chaff: mitigating the effects of noise in a plastome phylogenomic data set from *Pinus* L. (Pinaceae). BMC Evol Biol 12:100.

Peredo, E.L. King U.M. & D.H. Les. 2013. The Plastid Genome of *Najas flexilis*: Adaptation to Submersed Environments Is Accompanied by the Complete Loss of the NDH Complex in an Aquatic Angiosperm. *PloS ONE* 8. doi: 10.1371/journal.pone.0068591

Petersen, G. Seberg, O. Davis, J.I. & D.W. Stevenson. 2006. RNA editing and phylogenetic reconstruction in two monocot mitochondrial genes. Taxon, 55, 871–886.

Petersen, G. Seberg, O. Cuenca, A. Stevenson, D.W. Thadeo, M. Davis, J.I. Graham S. & T.G. Ross. 2016. Phylogeny of the Alismatales (Monocotyledons) and the relationship of *Acorus* (Acorales ?). Cladistics 32: 141–159.

Petersen, G. Cuenca, A. Zervas, A. Ross, G. T. Graham, S. W. Barrett, C. F. Davis J.I. & O. Seberg 2017. Mitochondrial genome evolution in Alismatales: Size reduction and extensive loss of ribosomal protein genes. *PLoS ONE, 12:* 1–21.

Philippe, H. Brinkmann, H. Lavrov, D.V, Littlewood, D.T.J. Manuel, M. Wörheide, G. & D. Baurain. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9, e1000602.

Rajan V. 2013. A method of alignment masking for refining the phylogenetic signal of multiple sequence alignments , Mol. Biol. Evol. 30: 689–712.

Rannala B. Zhu T. & Z. Yang. 2012. Tail paradox, partial identifiability, and influential priors in Bayesian branch length inference. Mol. Biol. Evol. 29: 325–335.

Rendle, A.B. 1899. A systematic revision of the genus *Najas*. Transactions of the Linnean Society of London: Botany 5: 379–436.

Ronquist, F. & J.P Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Ronquist, F. Teslenko, M. Van Der Mark, P. Ayres, D.L. Darling, A. Höhna, S. Larget, B. et al. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539–542.

Rosendahl, C.O. & Butters, F.K. 1935. The genus *Najas* in Minnesota. Rhodora 37: 345– 348.

Ross, T.G. Barrett, C.F. Soto, M. Lam, V.K.Y. Henriquez, C.L. Les, D.H. Davis, J.I. et al. 2016. Plastid phylogenomics and molecular evolution of Alismatales. Cladistics 32: 160–178.

Rudall, P.J, Chen, E.D. & E. Cullen. Evolution and development of monocot stomata. American Journal of Botany. 8: 1122–2241.

Song, Y. Yu, W.B. Tan, Y.H. Liu, B. Yao, X. Jin, J.J. Padmanaba, M. Yang, J.B. Corlett, R.T. 2016. Evolutionary comparisons of the chloroplast genome in Lauraceae and insights into loss events in the Magnoliids.Genome Biol Evol 9 (9):2354–2364.

Schwarz, G. 1978. Estimating the dimension of a model. The Annals of Statistics 6: 461–464.

Shaffer-Fehre, M. 1991a. The endotegmen tuberculae: an account of little-known structures from the seed coat of the Hydrocharitoideae (Hydrocharitaceae) and *Najas* (Najadaceae). Botanical Journal of the Linnean Society, 107: 169–188.

Shaffer-Fehre, M. 1991b. The position of *Najas* within the subclass Alismatidae (Monocotyledones) in the light of new evidence from seed coat structures in the Hydrocharitoideae (Hydrocharitales). Botanical Journal of the Linnean Society, 107: 189–209.

Sullivan, J. & P. Joyce. 2005. Model Selection in Phylogenetics. Annual Review of Ecology, Evolution, and Systematics 36: 445–466.

Sun, Y.X. Moore, M.J. Meng, A.P. Soltis, P.S. Soltis, D.E. Li, J.Q. & H.C. Wang. 2013. Complete plastid genome sequencing of Trochodendraceae reveals a significant expansion of the inverted repeat and suggests a Paleogene divergence between the two extant species.PLoS One. 8(4):e60429.

Tan, G. Muffato, M. Ledergerber, C. Herrero, J. Goldman, N. Gil, M. Dessimoz, C. 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. Syst. Biol. 1–14.

Talavera, G. & Castresana, J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56:564–577.

Tanaka, N. Setoguchi, H. & Murata, J. 1997. Phylogeny of the family Hydrocharitaceae inferred from rbcL and matK gene sequence data. Journal of Plant Research, 110: 329–33.

Tonti-Filippini, J. Nevill, P.G. Dixon, K. & I. Small. 2017. What can we do with 1000 plastid genomes? Plant J. 4: 808–818.

Triest, L. 1988. A revision of the genus *Najas* L. (Najadaceae) in the Old World. Mémoires de l'Académie Royale des Sciences d'Outre-Mer . Classe des Sciences Naturelles et Médicales 22: 1–172.

Wysocki, W.P. Clark, L.G. Attigala, L. Ruiz-Sanchez, E. Duvall, M.R. 2015. Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. BMC Evol Biol. 15:50.

Xie, W. Lewis, P.O. Fan, Y. Kuo, L. & M.H. Chen. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection, Syst. Biol. 60: 150–160.

Yan, L. Lai, X. Li, X. Wei, C. Tan, X. & Y. Zhang. 2015. Analyses of the complete genome and gene expression of chloroplast of sweet potato [*Ipomoea batata*]. PLoS One. 10(4), e0124083.

Zhang, C. Rannala, B. & Z. Yang. 2012. Robustness of compound Dirichlet priors for Bayesian inference of branch lengths. Syst. Biol.10 doi:10.1093/sysbio/sys030

Zhu, J. & X. Xu. 2015. The phylogeographic structure of *Hydrilla verticillata* (Hydrocharitaceae) in China and its implications for the biogeographic history of this worldwide-distributed submerged macrophyte. BMC Evolutionary Biology 15: 95.

Figure Legends

Figure 1. A) Cladogram representing 16 out of 17 Hydrocharitaceae genera previously represented in molecular analyses and outgroup *Butomus* (Butomaceae). The monotypic genus *Appertiella*, endemic to Madagascar, has not been sampled in any molecular analysis to date. Thick lines represent well supported clades, and stars represent long branches in previous analyses. Four subfamilies are recognized in Hydrocharitaceae, with all molecular analyses to date placing *Najas* within subfamily *Hydrilloideae*. **B)** Cladogram representing North American *Najas* taxa with plastomes sequenced in this study (Chapter 3) and used in the phylogenetic analysis here. Two accessions of *N. minor* [USA1 and USA2] are included, representing two separate introductions of this taxon into North America. These accessions were sequenced to contribute to another study (Les et al. 2015).

Figure 2. Phylogeny inferred from a concatenated dataset of 63 chloroplast protein-coding genes and four rRNA genes representing 16 genera of Hydrocharitaceae and ten *Najas* accessions. Support values for nodes with greater than 100% bootstrap support and 1.00 posterior probabilities have been removed. Remaining support values are from left to right IQTREE-UF bootstrap, 1000 standard non-parametric bootstraps, and Bayesian posterior probabilities. Branch lengths correspond to estimated number of substitutions per site.

Figure 3. Phylogenetic analysis of 63 chloroplast protein-coding genes and four rRNA genes representing ten *Najas* accessions, with inclusion of different outgroup taxa from Hydrocharitaceae and Butomaceae. Tree support values represent Bayesian posterior probabilities (with values for fully supported nodes removed). The position of subgenus *Najas* is unresolved with these data, variably associating with different taxon groups with poor support. Colors represent patterns obtained in maximum likelihood and Bayesian inference (left: ML, right: BI) with respect to the placement of subgenus *Najas*, either as sister to subgenus *Caulina* (orange), as sister to section *Euvaginatae* (blue) or as sister to section *Americanae* (green). Values in parentheses represent bootstrap values for corresponding pattern in ML analyses (1000 bootstraps). The top left tree is a mid-point rooted tree of *Najas*, based on the same data (1230 parsimony informative characters) with full clade support from the ML analysis also. Branch lengths correspond to estimated number of substitutions per site.

Figure 4. Four separate hydrocharit phylogenies inferred from chloroplast photosynthesis, plastid-encoded polymerase (PEP), ribosomal protein and ribosomal RNA genes (Ribosome) genes. Branch support values are 1000 non-parametric bootstraps (left) and Bayesian posterior probabilities (right). Values for fully supported branches have been removed. Branch lengths correspond to estimated number of substitutions per site.

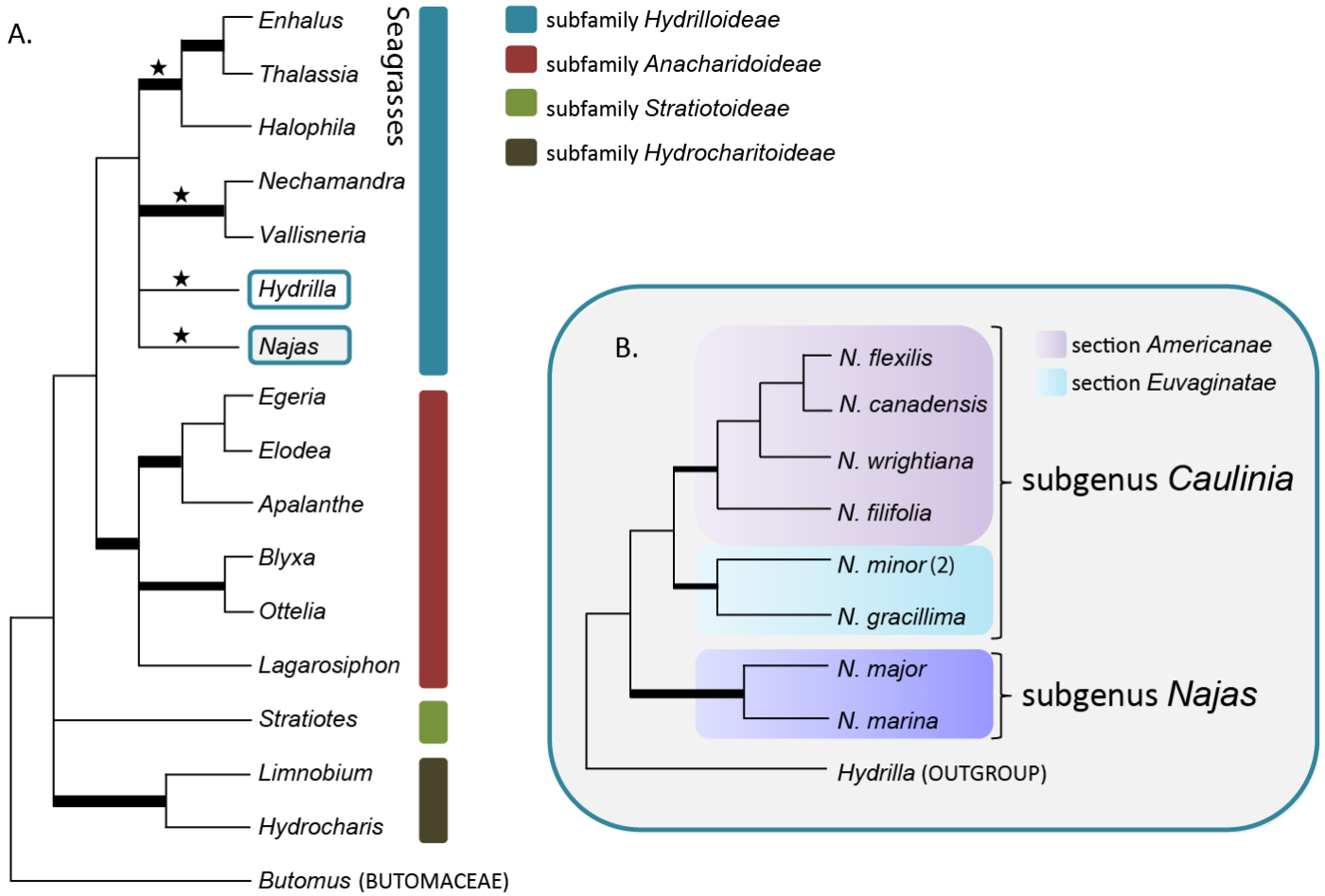

A.
Enhalus
Thalassia
Halophila
Nechamandra
Vallisneria
Hydrilla
Najas
Egeria
Elodea
Apalanthe
Blyxa
Ottelia
Lagarosiphon
Stratiotes
Limnobium
Hydrocharis
Butomus (BUTOMACEAE)
Seagrasses
subfamily Hydrilloideae
subfamily Anacharidoideae
subfamily Stratiotoideae
subfamily Hydrocharitoideae
B.
N. flexilis
N. canadensis
N. wrightiana
N. filifolia
N. minor (2)
N. gracillima
N. major
N. marina
Hydrilla (OUTGROUP)
section Americanae
section Euvaginatae
subgenus Caulinia
subgenus Najas

Figure 1

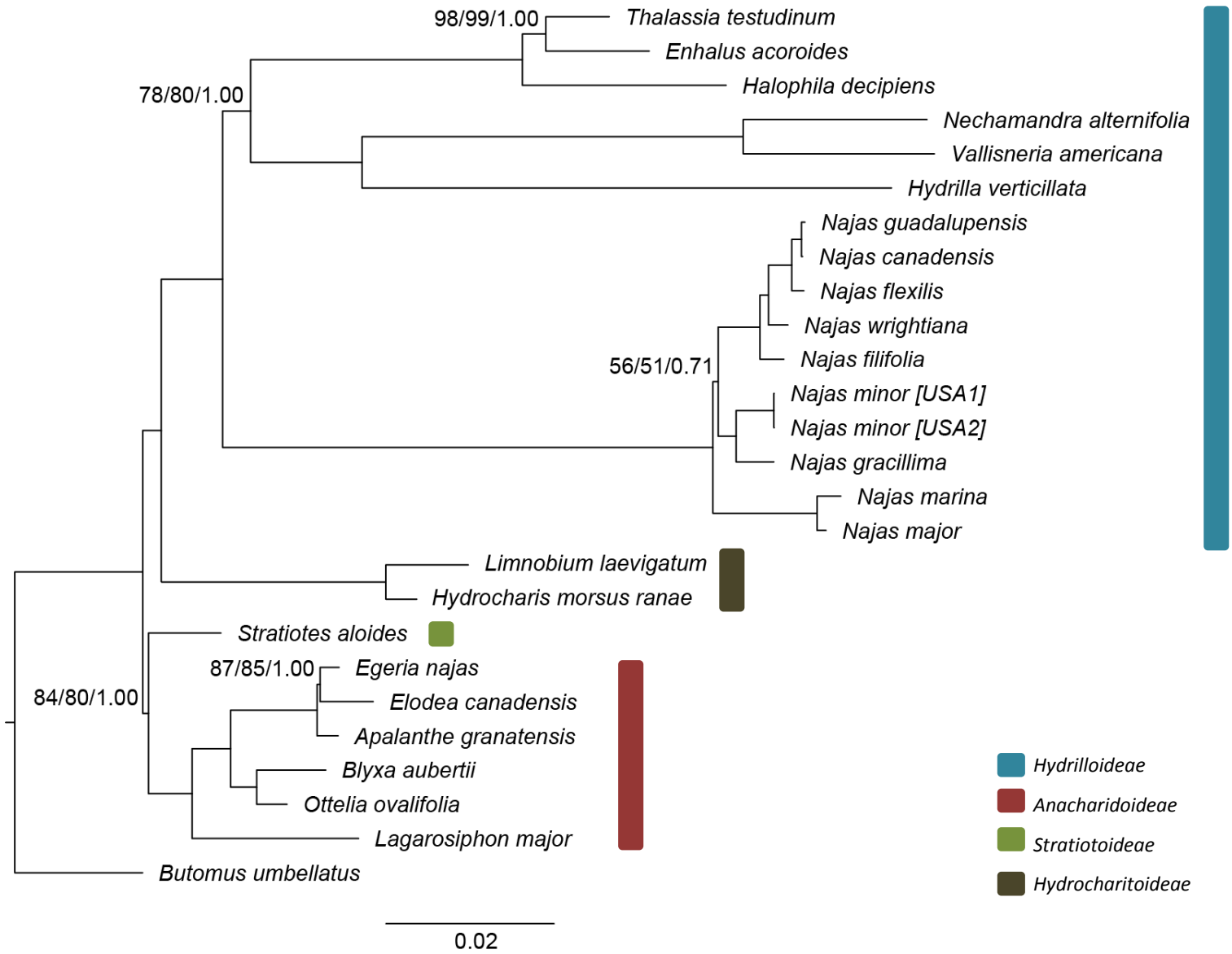

98/99/1.00
Thalassia testudinum
Enhalus acoroides
78/80/1.00
Halophila decipiens
Nechamandra alternifolia
Vallisneria americana
Hydrilla verticillata
Najas guadalupensis
Najas canadensis
Najas flexilis
Najas wrightiana
56/51/0.71
Najas filifolia
Najas minor [USA1]
Najas minor [USA2]
Najas gracillima
Najas marina
Najas major
Limnobium laevigatum
Hydrocharis morsus ranae
Stratiotes aloides
87/85/1.00
Egeria najas
84/80/1.00
Elodea canadensis
Apalanthe granatensis
Blyxa aubertii
Ottelia ovalifolia
Lagarosiphon major
Butomus umbellatus
0.02
Hydrilloideae
Anacharidoideae
Stratiotoideae
Hydrocharitoideae

Figure 2

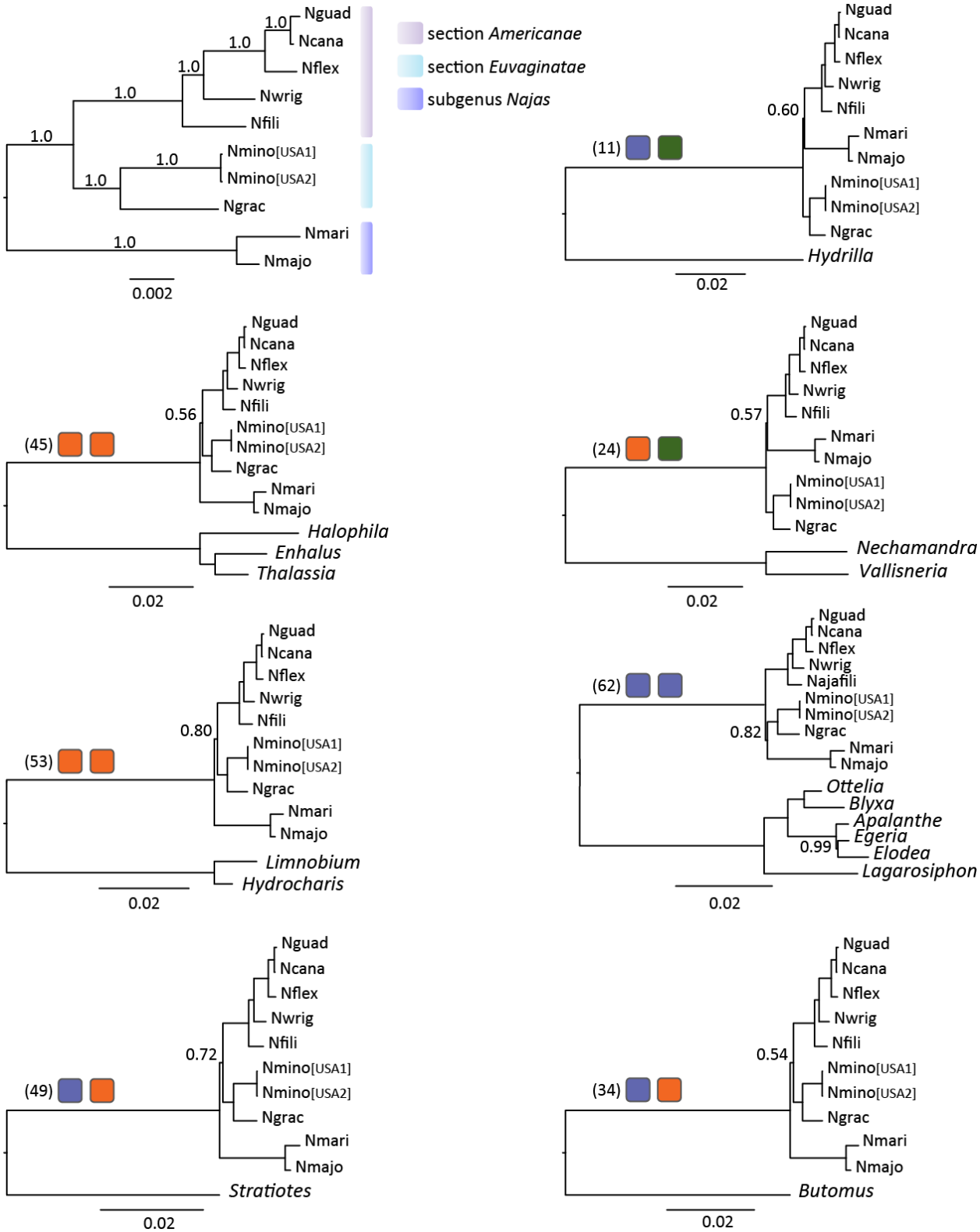

section Americanae
section Euvaginatae
subgenus Najas
Nguad
Ncana
Nflex
Nwrig
Nfili
Nmino[USA1]
Nmino[USA2]
Ngrac
Nmari
Nmajo
1.0
0.002
(11)
0.60
Hydrilla
0.02
(45)
0.56
Halophila
Enhalus
Thalassia
(24)
0.57
Nechamandra
Vallisneria
(53)
0.80
Limnobium
Hydrocharis
(62)
Najafili
0.82
Ottelia
Blyxa
Apalanthe
Egeria
0.99
Elodea
Lagarosiphon
(49)
0.72
Stratiotes
(34)
0.54
Butomus

Figure 3

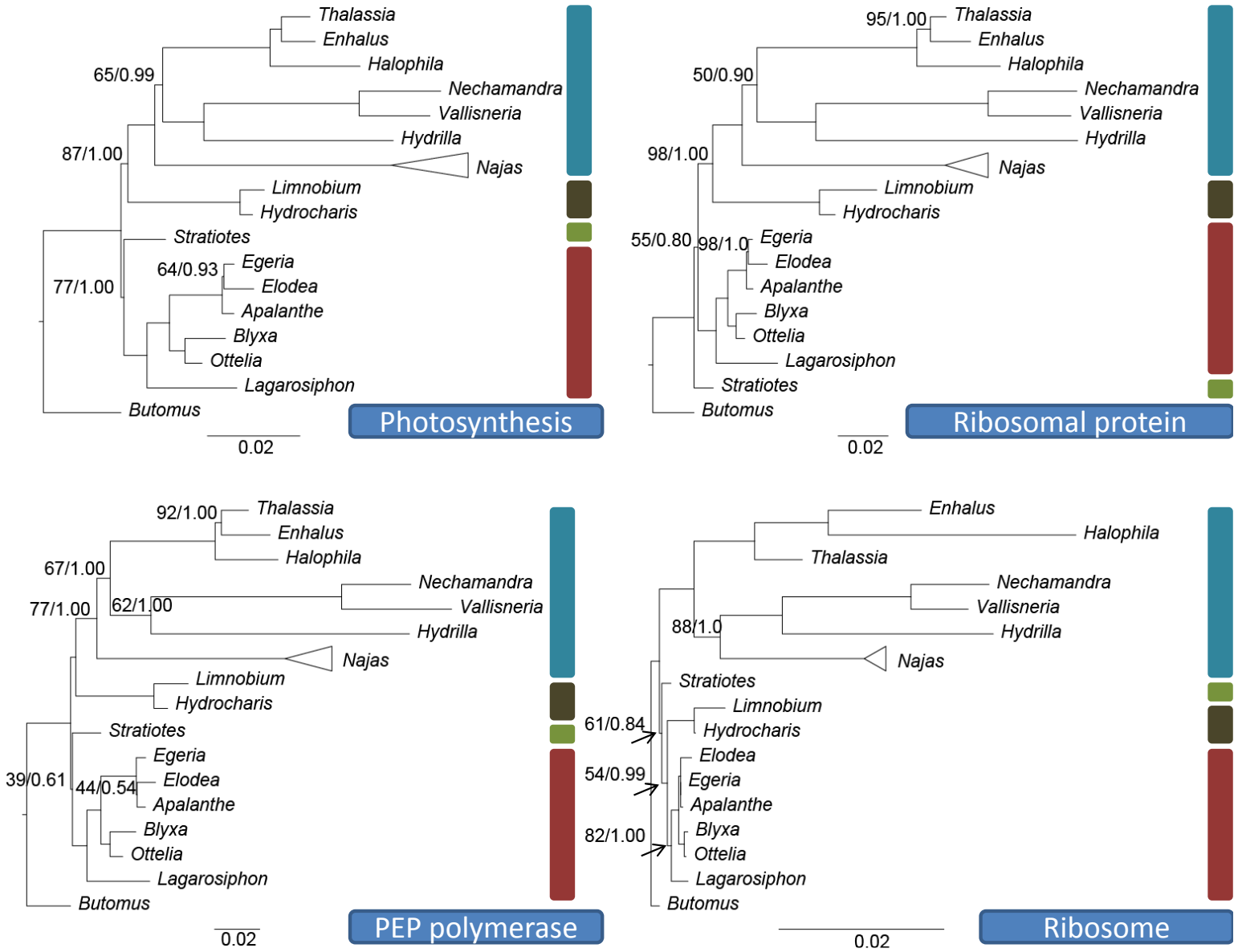

Thalassia
Enhalus
Halophila
65/0.99
Nechamandra
Vallisneria
Hydrilla
87/1.00
Najas
Limnobium
Hydrocharis
Stratiotes
64/0.93
Egeria
77/1.00
Elodea
Apalanthe
Blyxa
Ottelia
Lagarosiphon
Butomus
0.02
Photosynthesis
95/1.00
Thalassia
Enhalus
Halophila
50/0.90
Nechamandra
Vallisneria
Hydrilla
98/1.00
Najas
Limnobium
Hydrocharis
55/0.80
98/1.0
Egeria
Elodea
Apalanthe
Blyxa
Ottelia
Lagarosiphon
Stratiotes
Butomus
0.02
Ribosomal protein
92/1.00
Thalassia
Enhalus
Halophila
67/1.00
Nechamandra
77/1.00
62/1.00
Vallisneria
Hydrilla
Najas
Limnobium
Hydrocharis
Stratiotes
Egeria
39/0.61
44/0.54
Elodea
Apalanthe
Blyxa
Ottelia
Lagarosiphon
Butomus
0.02
PEP polymerase
Enhalus
Halophila
Thalassia
Nechamandra
Vallisneria
88/1.0
Hydrilla
Najas
Stratiotes
Limnobium
61/0.84
Hydrocharis
Elodea
54/0.99
Egeria
Apalanthe
Blyxa
82/1.00
Ottelia
Lagarosiphon
Butomus
0.02
Ribosome

Figure 4

Chapter 1. List of appendices

Appendix A. Accessions of *Najas guadalupensis* s.l. used in molecular analyses.

Appendix B. Gene region characteristics, primers and MrBayes parameters.

Appendix C. List of haplotypes and ribotypes amplified in *N. guadalupensis* s.l.

Appendix D. Chloroplast haplotypes in *Najas guadalupensis.*

Appendix E. nrITS universal primer variation in *Najas guadalupensis.*

Appendix F. Nflex-repeat specific primer variation in *Najas guadalupensis.*

Appendix G. Seed images in *Najas guadalupensis.*

Appendix H. Leaf images in *Najas guadalupensis.*

Appendix I. Overwintering bud/turion in *Najas guadalupensis*, Mansfield Hollow, Connecticut.

Appendix J. Taxa for seed and flower counts.

APPENDIX A

Lab_number	Species_designation	Country	State	County	Lake	Collector	Number	Date coll.	Latitude	Longitude	Herb.
Najaguad005	Najas guadalupensis	USA	CT	Litchfield Co.	Tyler Lake	D. Les & S. Sheldo	729	9/18/2007	41.836204	-73.260318	CONN
Najaguad102	Najas guadalupensis	USA	MN	Crow Wing Co.	Ruth Lake	D. Les	954	7/28/2010	46.749815	-93.953027	CONN
Najaguad105	Najas guadalupensis	USA	WI	Oconto	Berry Lake	D. Les	964	7/30/2010	44.888364	-88.477927	CONN
Najaguad106	Najas guadalupensis	USA	WI	Marquette Co.	Lake Montello	D. Les	968	7/31/2010	43.798062	-89.335623	CONN
Najaguad107	Najas guadalupensis	USA	WI	Jefferson Co.	Lake Ripley	D. Les	972	7/31/2010	43.005593	-88.988203	CONN
Najaguad113	Najas guadalupensis	USA	MI	Kalamazoo Co.	Jackson Hole Lake	D. Les	989	8/2/2010	42.313419	-85.356622	CONN
Najaguad117	Najas guadalupensis	USA	MN	Hennepin Co.	Snail Lake	D. Perleberg/S. Lc	821/1822	8/17/2010	45.07407	-93.126121	CONN
Najaguad118	Najas guadalupensis	USA	MN	Stearns	Julia Lake	D. Perleberg & S. .	824/1823	8/17/2010	47.667732	-94.887645	CONN
Najaguad119	Najas guadalupensis	USA	MN	Morrison	Crookneck Lake	D. Perleberg & S. .	820/1819	8/18/2010	46.244506	-94.611016	CONN
Najaguad120	Najas guadalupensis	USA	MN	Kandihoji Co.	Norway Lake	D. Perleberg & S. .	816/1815	8/17/2010	45.313889	-95.120000	CONN
Najaguad121	Najas guadalupensis	USA	MN	Anoka Co.	Linwood Lake	D. Perleberg & S. .	825/1826	8/17/2010	45.352592	-93.107209	CONN
Najaguad123	Najas guadalupensis	USA	CT	Tolland Co.	Lower Bolton Lake	N. Tippery	495	8/12/2010	41.80196	-72.431781	CONN
Najaguad125	Najas guadalupensis	USA	CT	Litchfield Co.	Bantam Lake	N. Murray	s.n.	8/31/2010	41.704417	-73.22206	CONN
Najaguad127	Najas guadalupensis	USA	CT	New London Co.	Gorton Pond	R. K. Shannon	s.n.	9/15/2010	41.342763	-72.21047	CONN
Najaguad136	Najas guadalupensis	USA	MN	Crow Wing Co.	Perch Lake	D. Les	946	7/28/2010	46.338046	-94.268283	CONN
Najaguad138	Najas guadalupensis	USA	CT	Tolland Co.	Mansfield Hollow State	U. King	556	10/22/2010	41.759315	-72.170438	CONN
Najaguad155	Najas guadalupensis	USA	WI	Walworth	Lulu Lake	T. Gerber	s.n.	8/8/2011	42.832651	-88.448866	CONN
Najaguad157	Najas guadalupensis	USA	WI	Florence Co.	Long Lake	P. Tikusis	s.n.	8/12/2011	45.841712	-88.673968	WI
Najaguad158	Najas guadalupensis	USA	WI	Forest	Pine Lake	E. Heath	s.n.	8/10/2011	45.676587	-88.979524	CONN
Najaguad179	Najas guadalupensis	USA	WA	King Co.	Steel Lake Park	D. Les	1039	7/31/2011	47.326125	-122.301671	CONN
Najaguad181	Najas guadalupensis	USA	WI	Walworth Co.	Whitewater Lake	N. Tippery	686	9/10/2011	42.745888	-88.71255	CONN
Najaguad193	Najas guadalupensis	USA	WI	Fond du Lac	Kettle Moraine Lake	C. Kolasinski	s.n.	9/20/2011	43.653558	-88.210814	CONN
Najaguad218	Najas guadalupensis	USA	PA	Butler Co.	Lake Arthur	D. Les	1088	7/5/2012	40.945885	-80.087571	CONN
Najaflor001	Najas floridana ??	USA	SC	Edgefield	Brick Pond Park	D. Les	752	6/23/2009	33.483668	-81.963135	CONN
Najaguad014	Najas guadalupensis	USA	NC	Stanly	Morrow Mountain Stat	D. Les	744	6/21/2009	35.363065	-80.073594	CONN
Najaguad015	Najas guadalupensis	USA	SC	Berkeley	Lake Moultrie	D. Les	747	6/23/2009	33.28402	-80.031144	CONN
Najaguad029	Najas guadalupensis	USA	CT	Tolland Co.	Lower Bolton Lake	N. Tippery	284	8/1/2009	41.801101	-72.433212	CONN
Najaguad030	Najas guadalupensis	USA	SC	Richland	Lake Wateree	K. Manuel	45	n/a	33.811283	-80.620917	CONN
Najaguad042	Najas guadalupensis	USA	MA	Berkshire Co.	Lake Onota	C. B. Hellquist	17,161	8/18/2009	42.468922	-73.280952	CONN
Najaguad047	Najas guadalupensis	USA	WI	Oneida Co.	Lake Minocqua	S. Knight	s.n.	5/3/2010	45.873862	-89.69477	CONN
Najaguad097	Najas guadalupensis	USA	MN	Kandihoji Co.	Norway Lake	D. Les	918	7/26/2010	45.304656	-95.108086	CONN

Najaguad099	Najas guadalupensis	USA	MN	Cass Co.	Horseshoe Lake	D. Les	939	7/27/2010	47.050804	-94.332298	CONN
Najaguad100	Najas guadalupensis	USA	MN	Crow Wing Co.	Perch Lake	D. Les	945	7/28/2010	46.339499	-94.266126	CONN
Najaguad101	Najas guadalupensis	USA	MN	Crow Wing Co.	Serpent Lake	D. Les	947	7/28/2010	46.48011	-93.927172	CONN
Najaguad103	Najas guadalupensis	USA	WI	Portage Co.	Lake Helen	D. Les	961	7/30/2010	44.618846	-89.242563	CONN
Najaguad104	Najas guadalupensis	USA	WI	Shawano Co.	Shawano Lake	D. Les	967	7/30/2010	44.807739	-88.524705	CONN
Najaguad108	Najas guadalupensis	USA	IL	Lake Co.	Round Lake	D. Les	975	8/1/2010	42.361298	-88.077572	CONN
Najaguad111	Najas guadalupensis	USA	IN	LaPorte	Silver Lake	D. Les	983	8/1/2010	41.691678	-86.59432	CONN
Najaguad114	Najas guadalupensis	USA	OH	Medina Co.	Leatha House Park	D. Les	1006	8/5/2010	41.114318	-82.049683	CONN
Najaguad049	Najas guadalupensis x flexilis	USA	CT	Litchfield Co.	Fisher Pond	N. Murray & G. Kr	s.n.	6/28/2010	42.035029	-73.424651	CONN
Najaguad149	Najas guadalupensis x flexilis	USA	CT	Litchfield Co.	Fisher Pond	A. M. Les	s.n.	6/11/2011	42.036172	-73.425705	CONN
Najaguad115	Najas guadalupensis x flexilis	USA	OH	Portage Co.	West Twin Lake	D. Les	1008	8/5/2010	41.197388	-81.338954	CONN
Najaguad109	Najas guadalupensis x flexilis	USA	IN	LaPorte	Stone Lake	D. Les	980	8/1/2010	41.611097	-86.748796	CONN
Najaguad110	Najas guadalupensis x flexilis	USA	IN	LaPorte	Stone Lake	D. Les	981	8/1/2010	41.611097	-86.748796	CONN
Najaguad116	Najas guadalupensis	USA	MN	Crow Wing Co.	Cross Lake	D. Perleberg & S.	s.n.	8/16/2010	46.662742	-94.117682	CONN
Najaguad152	Najas guadalupensis	USA	NY	St. Laurence	Black Lake	R. K. Shannon	1243	8/6/2011	44.505218	-75.603372	CONN
Najaguad153	Najas guadalupensis	USA	WI	Florence Co.	West Bass Lake	P. Tikusis	s.n.	8/11/2011	45.777309	-88.337614	WI
Najaguad154	Najas guadalupensis	USA	WI	Florence Co.	West Bass Lake	P. Tikusis	s.n.	8/11/2011	45.777309	-88.337614	WI
Najaguad156	Najas guadalupensis	USA	WI	Walworth	Lulu Lake	T. Gerber	s.n.	8/8/2011	42.832651	-88.448866	CONN
Najaguad160	Najas guadalupensis	USA	NY	Cattaraugus	Red House Lake	R. K. Shannon	1259	8/11/2011	42.103622	-78.746207	CONN
Najaguad163	Najas guadalupensis	USA	CT	Litchfield Co.	Bantam Lake	Murray & Hunt	2011-026	8/4/2011	41.703752	-73.224171	CONN
Najaguad164	Najas guadalupensis	USA	MN	Hennepin Co.	Christmas Lake	S. Simon & S. Eini	s.n.	8/24/2011	44.89617	-93.543136	CONN
Najaguad183	Najas guadalupensis	USA	WI	Oneida Co.	Lake Minocqua	S. Knight	s.n.	8/31/2011	45.873862	-89.69477	CONN
Najaguad184	Najas guadalupensis	USA	WI	Oneida Co.	Lake Tomahawk	S. Knight	s.n.	8/31/2011	45.817426	-89.60772	CONN
Najaguad185	Najas guadalupensis	USA	WI	Vilas Co.	Forest Lake	S. Knight	s.n.	9/2/2011	46.147682	-89.376377	CONN
Najaguad192	Najas guadalupensis	USA	CT	New London Co.	Rogers Lake	N. Murray (Collec	2011-034	8/15/2011	41.363659	-72.300032	CONN
Najaguad195	Najas guadalupensis	USA	WI	Adams	Goose Lake	R. Evans	s.n.	10/3/2010	43.798045	-89.637523	CONN
Najaguad201	Najas guadalupensis	USA	CT	New London Co.	Rogers Lake	A. Les	s.n.	7/21/2011	41.363659	-72.300032	CONN
Najaguad216	Najas guadalupensis	USA	PA	Indiana Co.	Yellow Creek Lake	D. Les	1080	7/5/2012	40.482484	-78.277228	CONN
Najaguad003	Najas guadalupensis x canade	USA	CT	Litchfield Co.	East Twin Lake	D. Les & S. Sheldo	725	9/18/2007	42.02149	-73.387044	CONN
Najaguad112	Najas guadalupensis x canade	USA	MI	Van Buren Co.	Maple Lake	D. Les	986	8/2/2010	42.232982	-85.88828	CONN
Najaguad217	Najas guadalupensis x canade	USA	PA	Indiana Co.	Yellow Creek Lake	D. Les	1084	7/5/2012	40.482484	-78.277228	CONN
Najaguad221	Najas guadalupensis	USA	OH	Ashtabula	Lake Roaming Rock	D. Les	1097	7/6/2012	41.64279	-80.829587	CONN
Najaguad224	Najas guadalupensis	USA	OH	Portage Co.	Michael J. Kirwan Reser	D. Les	1100	7/6/2012	41.15335	-81.083744	CONN
Najaguad225	Najas guadalupensis	USA	OH	Summit Co.	Turkeyfoot Lake	D. Les	1104	7/6/2012	40.964168	-81.538269	CONN

Najaguad226	Najas guadalupensis	USA	IN	Clark Co.	Deam Lake	D. Les	1118	7/13/2012	38.468771	-85.861415	CONN
Najaguad233	Najas guadalupensis	USA	WA	King Co.	Lake Kathleen, Renton,	K. Lanan	s.n.	2/8/2015	47.614450	-122.377148	CONN
Najaguad234	Najas guadalupensis	USA	WA	King Co.	Lake Kathleen	K. Lanan	s.n.	2/8/2015	47.614450	-122.377148	CONN
Najaguad243	Najas guadalupensis	USA	CT	Tolland Co.	Coventry Lake	U.King	s.n.	9/23/2015	41.765587	-72.309598	CONN
Najaguad256 (wa	Najas guadalupensis	USA	MI	Oakland Co.	Lake Orion, Lake 16	D.Les	742	8/9/2008	42.756941	-83.290945	CONN
Najaguad048	Najas guadalupensis	USA	WI	Vilas Co.	Fishtrap Lake	S. Knight	s.n.	5/15/2010	46.140271	-89.581748	CONN
											CONN
Najaguad006	Najas guadalupensis	USA	CT	Litchfield Co.	Bantam Lake	D. Les & S. Sheldo	719	9/18/2007	41.702202	-73.221981	CONN
Najaguad013	Najas guadalupensis	USA	NC	Guilford	Hagan Stone Park	D. Les	743	6/21/2009	35.951684	-79.733688	CONN
Najaguad037	Najas guadalupensis	USA	CT	Windham Co.	Black Pond	D. Les	807	8/26/2009	41.969403	-72.069037	CONN
Najaguad039	Najas guadalupensis	USA	CT	New Haven	West Lake	D. Les	815	9/24/2009	41.338215	-72.732597	CONN
Najaguad051	Najas guadalupensis	USA	CT	Litchfield Co.	Long Meadow Lake	D. Les	826	6/22/2010	41.651826	-73.208293	CONN
Najaguad142	Najas guadalupensis	USA	CT	Tolland Co.	Mansfield Hollow State	B. Capers	s.n.	1/9/2011	41.759315	-72.170438	CONN
Najaguad186	Najas guadalupensis	USA	CT	Litchfield Co.	Tyler Lake	H. Razifard	s.n.	9/10/2011	41.838408	-73.257673	CONN
Najaguad194	Najas guadalupensis	USA	NJ	Sussex	Tamarack Lake	R. K. Shannon	1315	9/30/2011	41.094803	-74.538642	CONN
Najaguad199	Najas guadalupensis	USA	NJ	Sussex	Wawayanda Lake	R. K. Shannon	1318	9/30/2011	41.181929	-74.432841	CONN
Najaguad206	Najas guadalupensis	USA	NH	Rockingham	Captain's Pond	E. Haug	s.n.	6/18/2012	42.808333	-71.172778	CDA
Najaguad248	Najas guadalupensis	USA	CT	Hartford Co.	Manitook Lake	R. Capers	509	7/15/2004	41.98545	-72.793620	CONN
Najaguad251	Najas guadalupensis	USA	CT	Litchfield Co.	Bantam Lake	Unknown	s.n.	9/14/1998	41.704722	-73.221667	CONN
Najaguad050	Najas guadalupensis	USA	CT	Litchfield Co.	Tom Pond	D. Les	825	6/22/2010	41.700389	-73.278748	CONN
Najaguad122	Najas guadalupensis	USA	CT	Windham Co.	Black Pond	N. Murray	2010-21	8/18/2010	41.970768	-72.071600	CONN
Najaguad137	Najas guadalupensis	USA	CT	Tolland Co.	Mansfield Hollow State	U. King	557	10/22/2010	41.759315	-72.170438	CONN
Najaguad002	Najas guadalupensis x canade	USA	CT	Litchfield Co.	Leonard Lake	D. Les & S. Sheldo	721	9/18/2007	41.706031	-73.474631	CONN
Najaguad004	Najas guadalupensis x canade	USA	CT	New Haven	Maltby Lake	S. Sheldon	s.n.	9/29/2007	41.307954	-72.978904	CONN
Najaguad045	Najas guadalupensis x canade	USA	CT	Litchfield Co.	Leonard Lake	S. Sheldon	s.n.	9/1/2008	41.706031	-73.474631	CONN
											CONN
Najaguad161/162	Najas olivacea??	USA	PA	Erie Co.	Edinboro Lake	R. K. Shannon	1260	8/14/2011	41.884968	-80.132711	CONN
Najaguad247	Najas olivacea??	USA	PA	Erie Co.	Edinboro Lake	R.K. Shannon	1571	12/26/2015	41.885000	-80.129000	CONN
Najaguad219	Najas olivacea??	USA	PA	Mercer Co.	Sandy Lake	D. Les	1089	7/6/2012	41.343443	-80.111267	CONN
Najaguad220	Najas olivacea??	USA	PA	Crawford Co.	Pymatuning Reservoir	D. Les	1092	7/6/2012	41.506969	-80.471167	CONN
Najaoliv098	Najas olivacea	USA	MN	Cass Co.	Townline Lake	D. Les	935	7/27/2010	47.069175	-94.222817	CONN
Najaoliv227	Najas olivacea	USA	WI	Sawyer Co.	Spider Lake	M. Berg	s.n.	n/a	46.09586	-91.120492	MICH
Najaoliv238	Najas olivacea (isolectotype)	USA	MN	Kandihoji Co.	Norway Lake	Rosendahl and Bu	6446	9/6/1933	45.314383	-95.107741	CONN
Najaoliv241	Najas olivacea	USA	MN	Cass Co.	Thunder Lake Township	D. Perleberg	s.n.	7/21/2009	46.983194	-94.015611	CONN

Najaoliv253	Najas olivacea	USA	MN	Cass Co.	Washburn Lake	D. Perleberg	s.n.	7/21/2009	46.875000	-93.991944	CONN
											CONN
Najaguad023	Najas guadalupensis	USA	MS	Pontotoc	Trace State Park	D. Les	791	7/5/2009	34.254751	-88.891011	CONN
Najaguad001	Najas guadalupensis	USA	CA	Butte	US Fish & Wildlife Servi	R. Whitkus	s.n.	8/19/2008	39.60408	-121.915615	CONN
Najaguad169	Najas guadalupensis	USA	CA	Butte	Gridley Game Refuge	D. Les	1051	8/9/2011	39.323969	-121.843033	CONN
Najaguad212	Najas guadalupensis	USA	CA	Imperial	11D recovery Pond at T	J. Johnson	s.n.	8/31/2001	32.996819	-115.532679	CDA
Najaguad214	Najas guadalupensis	USA	CA	El Dorado	Placerville, Spring Vale	W. West	s.n.	7/16/2009	38.75764	-120.93925	CDA
Najaguad031	Najas guadalupensis	USA	IA	Mills	Mile Hill Lake	N. Harms	s.n.	8/8/2009	41.045818	-95.784234	CONN
Najaguad071	Najas guadalupensis	USA	KS	Anderson	Garnett	D. Les	863	7/17/2010	38.295851	-95.240753	CONN
Najaguad093	Najas guadalupensis	USA	NE	Lincoln	Jeffrey Reservoir	D. Les	908	7/23/2010	40.94621	-100.410159	CONN
Najaguad135	Najas guadalupensis	USA	CA	Alameda	Lake Chabot	B. Ertter & D. Gro	20575	9/4/2010	37.730528	-122.120853	CONN
Najaguad170	Najas guadalupensis	USA	NM	Grant Co.	Bear Cannyon Lake	D. Les	1012	7/18/2011	32.883545	-107.996519	CONN
Najaguad175	Najas guadalupensis	USA	AZ	Maricopa	Saguaro Lake	D. Les	1022	7/22/2011	33.57031	-111.524472	CONN
Najaguad176	Najas guadalupensis	USA	CA	San Diego Co.	Lake Poway	D. Les	1025	7/23/2011	33.00669	-117.009202	CONN
Najaguad178	Najas guadalupensis	USA	CA	Alameda	Shadow Cliffs Regional	D. Les	1030	7/25/2011	37.670445	-121.840915	CONN
Najaguad063	Najas guadalupensis	USA	TX	Live Oak	Choke Canyon State Pa	D. Les	851	7/12/2010	28.473779	-98.247534	CONN
Najaguad072	Najas guadalupensis	USA	KS	Lynn	Mound City Lake	D. Les	865	7/17/2010	38.129347	-94.799827	CONN
Najaguad021	Najas guadalupensis	USA	MS	Forrest Co.	Geiger Lake Paul B. Joh	D. Les	787	7/3/2009	31.137361	-89.236778	CONN
Najaguad016	Najas guadalupensis	USA	GA	Butts	Indian Springs State Par	D. Les	754	6/24/2009	33.247993	-83.92844	CONN
Najaguad089	Najas guadalupensis	USA	IA	Warren	Lake Ahquabi	D. Les	900	7/22/2010	41.290273	-93.591633	CONN
Najaguad012	Najas guadalupensis	USA	TX	Denton	LAERF Pond	C. Owens	s.n.	6/9/2009	33.064111	-96.988333	CONN
Najaguad017	Najas guadalupensis	USA	AL	Chambers Co.	Lafayete City Lake	D. Les	755	6/24/2009	32.894797	-85.410334	CONN
Najaguad018	Najas guadalupensis	USA	GA	Decatur	Lake Seminole	D. Les	762	6/25/2009	30.749989	-84.847589	CONN
Najaguad020	Najas guadalupensis	USA	FL	Santa Rosa Co.	Milton: Public boat laur	D. Les	786	7/2/2009	30.631972	-87.03775	CONN
Najaguad022	Najas guadalupensis	USA	LA	Richland	Poverty Point State Par	D. Les	789	7/5/2009	32.490279	-91.491718	CONN
Najaguad032	Najas guadalupensis	USA	IA	Fremont	Percival Lake	N. Harms	s.n.	8/8/2009	40.750353	-95.820078	CONN
Najaguad043	Najas guadalupensis	USA	OK	Bryan	Durant State Fish Hatch	C. B. Hellquist	17168	9/3/2009	34.071997	-96.335057	CONN
Najaguad044	Najas guadalupensis	USA	OK	Love	Lake Murray	C. B. Hellquist	17169	9/3/2009	34.053266	-97.063289	CONN
Najaguad054	Najas guadalupensis	USA	IN	Jackson	Knob Lake	D. Les	831	7/8/2010	38.863735	-86.003543	CONN
Najaguad055	Najas guadalupensis	USA	IN	Orange	Patoka Lake	D. Les	832	7/8/2010	38.424518	-86.663627	CONN
Najaguad056	Najas guadalupensis	USA	AR	Mississippi	Big Lake National Wildli	D. Les	834	7/9/2010	35.873897	-90.098845	CONN

Najaguad057	Najas guadalupensis	USA	AR	Little River	Milwood Lake	D. Les	836	7/10/2010	33.723299	-93.988321	CONN
Najaguad058	Najas guadalupensis	USA	LA	Caddo Parish	Caddo Lake	D. Les	839	7/11/2010	32.710442	-94.018465	CONN
Najaguad059	Najas guadalupensis	USA	TX	Harrison	Brandy Branch Reservo	D. Les	841	7/11/2010	32.444328	-94.471654	CONN
Najaguad060	Najas guadalupensis	USA	TX	Bastrop	Bastrop Lake	D. Les	844	7/11/2010	30.154984	-97.284498	CONN
Najaguad061	Najas guadalupensis	USA	TX	Goliad	Coleto Lake Park	D. Les	846	7/12/2010	28.724259	-97.198438	CONN
Najaguad064	Najas guadalupensis	USA	TX	Uvalde	Nueces River	D. Les	853	7/13/2010	29.204455	-99.77679	CONN
Najaguad066	Najas guadalupensis	USA	TX	Hays	Dripping Springs	D. Les	856	7/13/2010	30.191278	-98.075783	CONN
Najaguad067	Najas guadalupensis	USA	OK	Murray	Lake of the Arbuckles	D. Les	857	7/14/2010	34.435517	-97.027439	CONN
Najaguad068	Najas guadalupensis	USA	OK	Comanche	Lake Elmer Thomas	D. Les	859	7/14/2010	34.724686	-98.520612	CONN
Najaguad069	Najas guadalupensis	USA	OK	Custer	Foss Lake	D. Les	860	7/15/2010	35.566718	-99.22867	CONN
Najaguad070	Najas guadalupensis	USA	KS	Trego	Cedar Bluff State Park	D. Les	861	7/16/2010	38.781741	-99.772055	CONN
Najaguad073	Najas guadalupensis	USA	KS	Wyandotte	Wyandotte County Parl	D. Les	867	7/18/2010	39.164386	-94.780061	CONN
Najaguad076	Najas guadalupensis	USA	MO	Callaway	Little Dixie Lake	D. Les	871	7/19/2010	38.909717	-92.124158	CONN
Najaguad077	Najas guadalupensis	USA	MO	Lincoln	Lake Tucci	D. Les	873	7/19/2010	38.84506	-91.043176	CONN
Najaguad078	Najas guadalupensis	USA	MO	Washington	Londell lake	D. Les	876	7/19/2010	38.203697	-90.819086	CONN
Najaguad079	Najas guadalupensis	USA	MO	St. Francois	Pim Lake	D. Les	877	7/19/2010	37.811893	-90.50075	CONN
Najaguad080	Najas guadalupensis	USA	MO	Madison Co.	SF Scout Ranch	D. Les	879	7/19/2010	37.639203	-90.329092	CONN
Najaguad082	Najas guadalupensis	USA	IL	Williamson	Devil's Kitchen Lake	D. Les	883	7/20/2010	37.641107	-89.102549	CONN
Najaguad083	Najas guadalupensis	USA	IL	Marion Co.	Forbes Lake	D. Les	886	7/20/2010	38.713147	-88.753018	CONN
Najaguad084	Najas guadalupensis	USA	IL	Effingham Co.	Lake Marion	D. Les	888	7/20/2010	39.123333	-88.618333	CONN
Najaguad085	Najas guadalupensis	USA	IL	Cole	Fox Ridge State Park, Ri	D. Les	889	7/21/2010	39.403564	-88.155271	CONN
Najaguad086	Najas guadalupensis	USA	IL	McLean	Moraine View State Par	D. Les	892	7/21/2010	40.409278	-88.725054	CONN
Najaguad087	Najas guadalupensis	USA	IA	Scott	Lake of the Hill Park	D. Les	896	7/22/2010	41.522658	-90.676083	CONN
Najaguad088	Najas guadalupensis	USA	IA	Mahaska	Lake Keomah State Parl	D. Les	897	7/22/2010	41.290858	-92.539185	CONN
Najaguad090	Najas guadalupensis	USA	IA	Madison Co.	Badger Creck State Park	D. Les	902	7/22/2010	41.476311	-93.918091	CONN
Najaguad091	Najas guadalupensis	USA	IA	Cass	Anita Lake	D. Les	904	7/22/2010	41.424359	-94.779787	CONN
Najaguad092	Najas guadalupensis	USA	NE	Hall	Grand Island L.E. Ray Pa	D. Les	906	7/23/2010	40.886178	-98.387578	CONN
Najaguad096	Najas guadalupensis	USA	NE	Madison Co.	Random Road Pond	D. Les	915	7/25/2010	42.039028	-97.441194	CONN
Najaguad148	Najas guadalupensis	USA	IL	Coles	Mattoon, Airpond Ponc	I. Klaus	8b(a)	7/28/2007	39.468804	-88.265868	CONN
Najaguad150	Najas guadalupensis	USA	MO	St. Charles Co.	Rotary Park Lake	D. Les	1011	7/11/2011	38.826943	-90.917592	CONN
Najaguad168	Najas guadalupensis	USA	CA	Yuba	Spenceville Wildlife Are	D. Les	1047	8/8/2011	39.089	-121.292801	CONN
Najaguad171	Najas guadalupensis	USA	AZ	Santa Cruz	Peña Blanca Lake	D. Les	1015	7/20/2011	31.407201	-111.084435	CONN
Najaguad172	Najas guadalupensis	USA	AZ	Pima	Arivaca Lake	D. Les	1017	7/20/2011	31.529524	-111.252656	CONN
Najaguad173	Najas guadalupensis	USA	AZ	Santa Cruz	Pategonia State Park	D. Les	1018	7/20/2011	31.494632	-110.853586	CONN

Najaguad174	Najas guadalupensis	USA	AZ	Maricopa	Papago Park	D. Les	1020	7/21/2011	33.453723	-111.947239	CONN
Najaguad223	Najas guadalupensis	USA	KY	Jefferson Co.	Fisherman's Park	D. Les	1117	7/13/2012	38.167635	-85.511865	CONN
Najaguad228	Najas guadalupensis	USA	WV	Upshur Co.	Stonecrop Lake	R. K. Shannon	1359	8/21/2012	38.962000	-80.317000	CONN
Najaguad229	Najas guadalupensis	USA	AZ	Yavapai Co.	Page Springs	D. Les	800	8/1/2009	34.767000	-111.894026	CONN
Najaguad235	Najas guadalupensis	USA	WY	Teton Co.	Kelly warm Springs	C.E. & C.B. Hellqu	1168-14	8/3/2014	43.639380	-110.615000	CONN
Najaguad236	Najas guadalupensis	USA	WV	Mingo Co.	Laurel Lake Wildlife Ma	R. K. Shannon	1506	10/26/2013	37.844000	-82.213000	CONN
Najaguad237	Najas guadalupensis	USA	KS	Douglas Co.	Douglas Lake	R. K. Shannon	1538	7/16/2014	38.801000	-95.163000	CONN
Najaguad245	Najas guadalupensis	USA	IL	Coles Co.	Matton: Airport pond	I.Klaus	8b(b)	8/28/2007	39.470278	-88.273611	CONN
Najaguad008	Najas guadalupensis	USA	LA	Denton Co.	LSU Aquaculture Resea	Urbatsch	s.n.	11/27/2007	30.371502	-91.182023	CONN
Najaguad145	Najas guadalupensis	USA	TX	Denton	LAERF Pond	C. Owens	s.n.	7/14/2009	33.046111	-96.994167	CONN
Najaguad033	Najas guadalupensis	USA	MO	Linn	Jo Shelby Lake	N. Harms	s.n.	8/13/2009	39.792962	-92.230087	CONN
Najaguad075	Najas guadalupensis	USA	MO	Columbia	Stephens Lake Park	D. Les	870	7/18/2010	38.95102	-92.307337	CONN
Najaguad010	Najas floridana??	USA	FL	Lee	Bonita Springs	J. Kunzer	s.n.	12/7/2007	26.339806	-81.778697	CONN
Najaflor002	Najas floridana	USA	FL	Volusia	Juniper Wayside, Ocala	D. Les	768	6/27/2009	29.213075	-81.654694	CONN
Najaflor003	Najas floridana	USA	FL	Marion Co.	Nelson's Outdoor Reso	D. Les	769	6/27/2009	28.991286	-81.835215	CONN
Najaflor006	Najas floridana	USA	FL	Lee	Pond 11, Lee Co Mosqu	D. Les	774	6/30/2009	26.646625	-81.706467	CONN
Najaflor007	Najas floridana	USA	FL	Lee	Pond 4, Lee Co Mosquit	D. Les	775	6/30/2009	26.648931	-81.712286	CONN
Najaflor009	Najas floridana	USA	FL	Lee	Veterans Memorial Hw	D. Les	781	6/30/2009	26.612711	-82.036658	CONN
Najaflor010	Najas floridana	USA	FL	Columbia	Ichetucknee Springs Sta	L. Benoit	67B	12/31/2009	29.963004	-82.767809	CONN
Najaflor014	Najas floridana	USA	FL	Collier	Florida Panther Wildlife	D. Les	773	6/27/2009	26.152126	-81.354626	CONN
Najaflor015 (was	Najas floridana	USA	FL	Miami	Florida International Ur	J. Fourqurean & J.	s.n.	11/16/2007	25.776306	-80.340127	CONN
Najaflor016 (was	Najas floridana	USA	FL	Hernando	Weeki Wachee River	C. B. Hellquist	17238	4/10/2012	28.531328	-82.615715	CONN
Najaguad019	Najas floridana??	USA	FL	Okaloosa	Wright Compost Recycl	D. Les	783	7/2/2009	30.47077	-86.635061	CONN
Najaguad132	Najas guadalupensis	USA	TX	Jeff Davis	Madera Creek	Barre Hellquist	17192	9/16/2010	30.693811	-104.123568	CONN
Najaguad133/134	Najas guadalupensis	USA	TX	Brewster	Elephant Mountain	C. Williams	s.n.	9/22/2010	30.034097	-103.531981	CONN
Najaguad074	Najas guadalupensis	USA	KS	Atchison	Atchison State Fishing L		868	7/18/2010	39.634321	-95.177274	CONN
Najaguad062	Najas guadalupensis	USA	TX	San Patricio	Welder Wildlife Refuge	D. Les	848	7/12/2010	28.122932	-97.368333	CONN

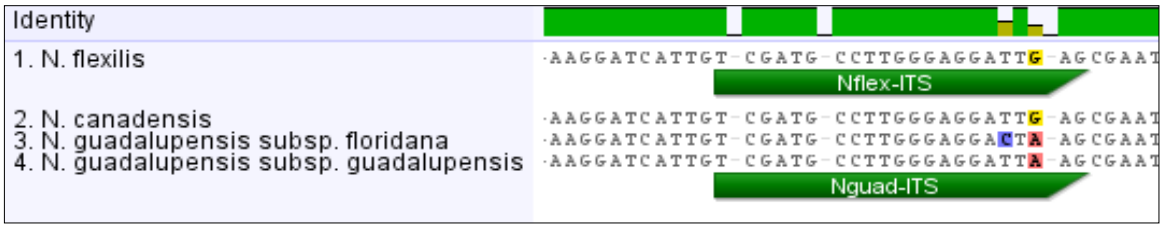
Najaguad065	Najas guadalupensis	USA	TX	Uvalde Co.	Garner State Park. Frio	D. Les	855	7/13/2010	29.576822	-99.730763	CONN
Najaguad124	Najas guadalupensis	USA	TX	Brewster	Elephant Mountain	C. Williams	s.n.	8/23/2010	30.057256	-103.493167	CONN
Najaguad177	Najas guadalupensis	USA	CA	San Diego	Dixon Lake	D. Les	1026	7/23/2011	33.160193	-117.043365	CONN
Najaguad094	Najas guadalupensis	USA	NE	Garden Co.	Islands Lake	D. Les	911	7/24/2010	41.72839	-102.399971	CONN
Najaguad095	Najas guadalupensis	USA	NE	Cherry Co.	Cottonwoon Lake State	D. Les	913	7/24/2010	42.914489	-101.675572	CONN
Najaguad196/197	Najas guadalupensis	USA	SD	Lyman	Lower Brule Indian Res	G. E. Larson	s.n.	8/8/2011	43.964444	-99.563333	CONN
Najaguad197	Najas guadalupensis	USA	SD	Lyman	Lower Brule Indian Res	G. E. Larson	s.n.	8/8/2011	43.964444	-99.563333	CONN
Najaguad244	Najas guadalupensis	USA	SD	Brookings Co.	Borrow pond, south of	Gary E. Larson	s.n.	7/22/2011	44.276389	-96.786111	CONN
Najaguad255 (wa	Najas guadalupensis	Hondura	Depar	5.5 km south of tov	Common in small pools	Philbrick & Rame	6276	4/12/2008	13.797660	-87.127250	CONN
Najaflor004	Najas floridana	USA	FL	Orange	Kraft Azalea Garden	D. Les	770	6/27/2009	28.611083	-81.335467	CONN
Najaflor005	Najas floridana	USA	FL	Orange	Lake Osceola	D. Les	772	6/27/2009	28.604615	-81.341319	CONN
Najaflor008	Najas floridana	USA	FL	Lee	County Cape Coral	D. Les	776	6/30/2009	26.674714	-81.933486	CONN
Najaguad011	Najas guadalupensis	USA	MN	Crow Wing Co.	Nisswa Lake	D. Perleberg	s.n.	6/15/2008	46.518891	-94.295723	CONN
Najaguad036	Najas guadalupensis	USA	CT	Tolland Co.	Mansfield Hollow State	D. Les	809	9/14/2009	41.757927	-72.169209	CONN
Najaguad126	Najas guadalupensis	USA	CT	Windham Co.	Roger Lake	R. K. Shannon	s.n.	9/15/2010	41.726148	-71.853814	CONN
Najaguad151	Najas guadalupensis	USA	CT	Hartford Co.	Barkhampstead Reserv	A. M. Les	s.n.	7/12/2011	42.00499	-72.94807	CONN
Najaguad159	Najas guadalupensis	USA	WI	Forest	Pine Lake	E. Heath	s.n.	8/10/2011	45.676587	-88.979524	CONN
Najaguad182	Najas olivacea??	USA	WI	Vilas Co.	Fishtrap Lake	S. Knight	s.n.	9/2/2011 herb	46.140271	-89.581748	CONN
Najaguad187	Najas guadalupensis	USA	CT	Tolland Co.	Bolton lake	H. Razifard	Sept 2011	9/7/2011 fron	41.805881	-72.431471	CONN
Najaguad189	Najas guadalupensis	USA	MA	Worcester Co.	Walker Pond	D. Les	1057	9/17/2011	42.13594	-72.060454	CONN
Najaguad190	Najas guadalupensis	USA	MA	Worcester Co.	Demond Pond	D. Les	1059	9/17/2011	42.351444	-71.972013	CONN
Najaguad191	Najas guadalupensis	USA	MA	Barnstable	Coonamessett Pond	E. L. Peredo	s.n.	9/25/2011	41.625088	-70.563857	CONN
Najaguad204	Najas guadalupensis	USA	VT	Chittenden	Lake Champlain, Mallet	S. Sheldon	2011 008	8/17/2011	44.5525	-73.207158	CONN
Najaguad205	Najas guadalupensis	USA	VT	Rutland Co.	Sunrise Lake	S. Sheldon	2011 001	n/a	43.760558	-73.260734	CONN
Najaguad242	Najas guadalupensis	USA	MD	Allegany Co.	Habbeb Lake	D. Les	828	7/7/2010	39.702743	-78.653821	CONN
Najaguad180	Najas guadalupensis	USA	PA	Wayne Co.	Lake Ladore	A. F. Rhoads & T.	s.n.	8/23/2005	41.571396	-75.389381	MOAR
Najaoliv232	Najas olivacea (lectotype)	USA	MN	Kandihoji Co.	Norway Lake	Rosendahl and Bu	6446	9/6/1933	45.314383	-95.107741	MIN
Najaoliv240	Najas olivacea	USA	MN	Crow Wing Co.	Mile Lake	S. Loso	s.n.	9/10/2008	46.344444	-94.328889	CONN
Najaoliv230	Najas olivacea	USA	WI	Bayfield Co.	Upper Eau, Claire Lake	M S. Berg	WGS-84	n/a	46.305170	-91.496510	CONN
Najaguad146	Najas guadalupensis	Costa Ri	Guana	Parque Nacional Palo Verde		G.E. Crow	7447	10/21/1989	10.37836	-85.336089	CONN
Najaguad081	Najas guadalupensis	USA	IL	Jackson	Murphrysboro State Pa	D. Les	882	7/20/2010	37.778628	-89.378102	CONN

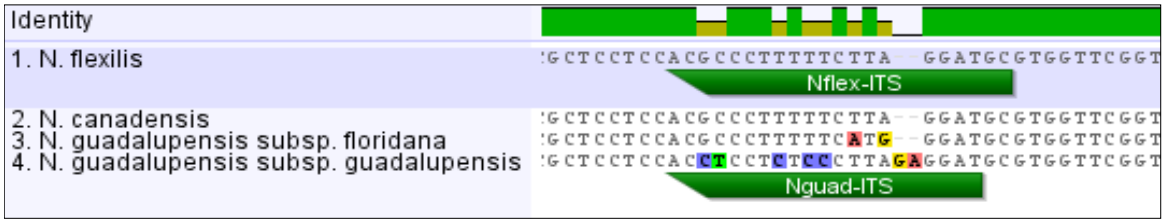
Najaguad024	Najas guadalupensis	USA	AL	Marshall	Guntersville Lake	D. Les	793	7/6/2009	34.509306	-86.159915	CONN
Najaguad147	Najas guadalupensis	USA	LA	Ochaita	Cheniere Lake	P. Thomas	125,134a	9/9/1991	32.453244	-92.191968	CONN
Najaguad198	Najas guadalupensis	USA	SD	Brookings	Borrow pond, south of	G. E. Larson	s.n.	7/24/2011	44.308446	-96.819638	CONN
Najaguad202	Najas guadalupensis	USA	TX	Collin	Lake Forest	C. B. Hellquist	17241	4/17/2012	33.195928	-96.678251	CONN
Najaguad207	Najas guadalupensis	USA	CA	Shasta Co.	Knighten Rd. (Anderson	B. Kreps	s.n.	9/19/2000	40.468304	-122.275888	CDA
Najaguad208	Najas guadalupensis	USA	CA	Riverside	Rancho de los Coyotes,	R. W. Shaffer	s.n.	4/16/2001	33.761731	-116.272457	CDA
Najaguad209	Najas guadalupensis	USA	CA	Stanislaus	Lake side apartments T	T. Palmer	s.n.	7/9/1979	37.485735	-120.846109	CDA
Najaguad210	Najas guadalupensis	USA	CA	Yuba	Reneissance Winery, 23	Finley	s.n.	9/1/2000	39.326848	-121.246934	CDA
Najaguad211	Najas guadalupensis	USA	CA	Alameda Co.	Shadow Cliff Lake	J. K. Wilson	s.n.	10/10/2010	37.668696	-121.837332	CDA
Najaguad213	Najas guadalupensis	USA	CA	Kern	Paradise Lake	J. Sithole	s.n.	6/15/2005	35.175249	-118.927078	CDA
Najaguad222	Najas guadalupensis	USA	WV	Marion Co.	Prickett's Fort State Par	D. Les	1122	7/14/2012	39.517306	-80.092222	CONN
Najaguad028	Najas guadalupensis	USA	CO	Denver	Rocky Mountain Lake	N. Tippery	276	7/23/2009	39.782356	-105.030132	CONN
Najaguad034	Najas guadalupensis	USA	MO	Linn	Fountain Grove Wildlife	N. Harms	s.n.	8/13/2009	39.715963	-93.318077	CONN
Najaguad040	Najas canadensis x guadalupei	USA	CT	New Haven	Maltby Lake	D. Les	817	9/24/2009	41.307961	-72.978893	CONN
Najaguad041	Najas canadensis x guadalupei	USA	CT	Litchfield Co.	East Twin Lake	D. Les	820	9/26/2009	42.026448	-73.387079	CONN
Najaguad046	Najas canadensis x guadalupei	USA	CT	Litchfield Co.	East Twin Lake	S. Sheldon	s.n.	9/1/2008	42.02149	-73.387044	CONN
Najaguad188	Najas canadensis x guadalupei	USA	CT	Litchfield Co.	East Twin Lake	D. Les	1069	9/21/2011	42.024316	-73.383071	CONN
	Outgroups										
Najawrig003	Najas wrightiana	USA	FL	Lee	Wild Turkey Strand Pre	S. Furnari	s.n.	rec 10-1005	26.566424	-81.689384	CONN
Najafili002	Najas filifolia	USA	FL	Santa Rosa	Public Boat Lake	D. Les	784		30.632106	-87.037292	CONN

APPENDIX B

Abbreviation	Primer region	Alignment length	Primer name	Primer sequences 5' - 3'	Notes
rbcL	rbcL gene	1138	1F	ATGTCACCACAAACAGAAACTAAAGC	
			1204R	CCTAAGGGTGTCCTAAAGTTTCTCC	
trnK/matK	3' matK gene	430	DL19	AGTACTCGGCTTTTAAGTGC	
	tRNA-K intron	723	matK_1255R	GATTTCAAGATGAATAGGATAGGG	
Total		**1153**			
trnL-F	tRNA-L intron	730	c primer	CGAAATCGGTAGACGCTACG	
	tRNA-L 5'	50	f primer	ATTTGAACTGGTGACACGAG	
	tRNA L-F intergenic	493			
	tRNA-F 3'	15		universal c and f primers of Taberlet et al. (1991)	
Total		**1288**		(138nts unalignable region removed from tRNA-L intron)	
Total chloroplast		**1539**			
Universal nrITS	18S gene	32	ITS5	GGAAGTAAAAGTCGTAACAAGG	
	ITS-1	284	ITS4	TCCTCCGCTTATTGATATGC	
	5.8S gene	160			
	ITS-2	241			
	26S gene	31			
Total		**748**			
Repeat-specific (Nflex-ITS)			Nflex_ITS_33F	TCGATGCCTTGGGAGGATTGAG	
			Nflex_ITS_3R	GCATCCTAAGAAAAAGGGCGT	
Repeat-specific (Nguad-ITS)			Nguad_ITS_3F	TCGATGCCTTGGGAGGATTAAG	
			Nguad_ITS_3R	ATCCTCTAAGGGAGAGGAGGT	
Splitstree	ITS-1	264			
	5.8S gene	160			
	ITS-2	14			
Total		**438**			

Nflex-ITS and Nguad-ITS repeat specific primers





Mr Bayes - Chloroplast haplotypes

```
    unlink statefreq=(all) revmat=(all) shape=(all) pinvar=(all); [this is unlinking priors]
    prset applyto=(all) ratepr=variable brlenspr=unconstrained:gammadir(1.0,0.100,1.0,1.0);


    lset applyto=(7) coding=variable;   [for the standard data...ascertainment bias]

    lset applyto=(1,2,5,6) nst=1;
    lset applyto=(3,4) nst=2;

    lset applyto=(1,3,4) rates =propinv;
    lset applyto=(2) rates =invgamma;
    lset applyto=(5,6) rates =equal;

    prset applyto=(3,4) tratiopr=Beta(1.0,1.0);

    prset applyto=(2)          shapepr=exp(1.0);
    prset applyto=(1,2,3,4)    pinvarpr=uniform(0.0,1.0);

    prset applyto=(1,5)      statefreqpr=fixed (equal); [this is placing fixed stationary state frequencies on partitions 2 and 3]
    prset applyto=(2,3,4,6) statefreqpr=Dirichlet (1.00,1.00,1.00,1.00);

    outgroup Najawrig003;
    mcmcp ngen= 10000000 relburnin=yes burninfrac=0.25 printfreq=1000  samplefreq=1000 nruns=2 nchains=4 savebrlens=yes filename = mcmc;
    mcmc;
    sumt;
END;

[my partitionfinder partition scheme chosen - by codon]
[Subset | Best Model |  Subset Sites / subs classes(rate parameters) / substitution rate priors  / rates across sites  / gamma shape prior  /prop. inv sites prior          / stationary state freq priors                ]
[1      | JC+I       | 0    | 524                 nst = 1                                        rates=propinv                                  pinvarpr=uniform(0.0,1.0)       statefreqpr=fixed (equal)                     ]
[2      | F81+I+G    | 0    | 522                 nst = 1                                        rates=invgamma        shapepr=exp(1.0)         pinvarpr=uniform(0.0,1.0)       statefreqpr=Dirichlet (1.00,1.00,1.00,1.00)   ]
[3      | HKY+I      | 0    | 522                 nst = 2              tratiopr=Beta(1.0,1.0)    rates=propinv                                  pinvarpr=uniform(0.0,1.0)       statefreqpr=Dirichlet (1.00,1.00,1.00,1.00)   ]
[4      | HKY+I      | 0    | 1453                nst = 2              tratiopr=Beta(1.0,1.0)    rates=propinv                                  pinvarpr=uniform(0.0,1.0)       statefreqpr=Dirichlet (1.00,1.00,1.00,1.00    ]
[5      | JC         | 0    | 65                  nst = 1                                        rates=equal                                                                    statefreqpr=fixed (equal)                     ]
[6      | F81        | 0    | 493                 nst = 1                                        rates=equal                                                                    statefreqpr=Dirichlet (1.00,1.00,1.00,1.00)   ]
[7      | (indels)   | 0    | 45                  nst = 1                                        rates=equal                                                                    statefreqpr=fixed (equal)                     ]
```

Mr Bayes - nrITS

```
    unlink statefreq=(all) revmat=(all) shape=(all) pinvar=(all);              [this is unlinking priors]
    prset applyto=(all) ratepr=variable brlenspr=unconstrained:gammadir(1.0,0.100,1.0,1.0);

    lset applyto=(1,3) nst=1;
    lset applyto=(2) nst=2;
    lset applyto=(1,2,3) rates=equal;

    lset applyto=(3) coding=variable;   [for the standard data...ascertainment bias]

    prset applyto=(2)tratiopr=beta(1.0,1.0);

    prset applyto=(1,3) statefreqpr=fixed (equal);       [this is placing fixed/equal stationary state frequencies on partitions 2 and 3]
    prset applyto=(2) statefreqpr=dirichlet(1.0,1.0,1.0,1.0);

    outgroup Najawrig003;
    mcmcp ngen= 10000000 relburnin=yes burninfrac=0.25 printfreq=1000  samplefreq=1000 nruns=2 nchains=4 savebrlens=yes filename=mcmc;
    mcmc;
    sumt;
END;
[My partitionfinder best partitions - Scheme = by_gene_spacer]

[Subset | Best Model |  Subset Sites / subs classes(rate parameters) / substitution rate priors  / rates across sites  / gamma shape prior  /prop. inv sites prior          / stationary state freq priors                ]
[1      | JC         | 0    | 556                 nst = 1                                        rates=equal              none                   none                    statefreqpr=fixed (equal)                     ]
[2      | HKY        | 0    | 192                 nst = 2              tratiopr=Beta(1.0,1.0)    rates=equal              none                   none                    statefreqpr=Dirichlet (1.00,1.00,1.00,1.00)   ]
[3      | (indels)   | 0    | 12                  nst = 1                                        rates=equal                                                             statefreqpr=fixed (equal)                     ]
```

APPENDIX C

dash = PCR amplified Nflex repeat primer with sequenced variant indicated in column

Yes = PCR gel band for both Nflex and Nguad repeat primers but not sequenced

n/a=PCR failed for both Nflex and Nguad repeat primers (or DNA not available or Low quality DNA or herbarium extractions)

No = PCR amplified from guad primer but not flex

	Chloroplast				Nflex-ITS repeat primer				Clones	Universal ITS primers plus Nguad repeat-specific primers															
Lab_number	**rbcL**	**trnK**	**trnLF**	**Haplo**	**nrUn**	**nrC**	**nrF**	**nrG1**		**nrC**	**nrF**	**nrG1**	**nrG2**	**nrG3**	**nrG4**	**nrG5**	**nrG6**	**nrG7**	**nrG8**	**nrG9**	**nrG10**	**nrG11**	**nrG12**	**nrG13**	**nrG14**
Najaguad005	n/a	trmkG8	n/a	cG16	n/a	n/a	n/a	n/a															nrG12		
Najaguad102	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad105	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad106	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad107	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad113	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad117	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad118	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad119	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad120	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad121	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad123	n/a	trmkG8	n/a	cG16	No	No	No	No															nrG12	nrG13	
Najaguad125	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12	nrG13	
Najaguad127	n/a	trmkG8	n/a	cG16	Faint	Faint	Faint	Faint															nrG12		
Najaguad136	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad138	n/a	trmkG8	n/a	cG16	No	No	No	No															nrG12		
Najaguad155	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad157	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad158	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad179	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad181	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad193	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad218	n/a	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		
Najaflor001	rbG2	trmkG8	n/a	cG16	-	-	-	nrG1	*			nrG1												nrG13	
Najaguad014	rbG2	trmkG8	n/a	cG16	No	No	No	No	*													nrG11	nrG12		
Najaguad015	rbG2	trmkG8	n/a	cG16	-	-	-	nrG1	*			nrG1									nrG10		nrG12		
Najaguad029	rbG2	trmkG8	n/a	cG16	No	No	No	No															nrG12		

Najaguad030	rbG2	trmkG8	n/a	cG16	No	No	No	No	*											nrG10		nrG12		
Najaguad042	rbG2	trmkG8	n/a	cG16	Faint	Faint	Faint	Faint														nrG12		
Najaguad047	rbG2	trmkG8	trnG8	cG16	-	-	nrF(p)	-														nrG12		nrG14
Najaguad097	rbG2	trmkG8	n/a	cG16	-	-	nrF(p)	-	*													nrG12		nrG14
Najaguad099	rbG2	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes														nrG12		nrG14
Najaguad100	rbG2	trmkG8	n/a	cG16	-	-	nrF	-														nrG12		nrG14
Najaguad101	rbG2	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes														nrG12		nrG14
Najaguad103	rbG2	trmkG8	n/a	cG16	-	-	nrF(p)	-														nrG12		nrG14
Najaguad104	rbG2	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes														nrG12		nrG14
Najaguad108	rbG2	trmkG8	n/a	cG16	-	-	nrF(p)	-														nrG12		nrG14
Najaguad111	rbG2	trmkG8	trnG8	cG16	-	nrC	nrF	-												nrG10		nrG12		
Najaguad114	rbG2	trmkG8	trnG8	cG16	-	nrC	-	-														nrG12		
Najaguad049	rbG2	trmkG8	n/a	cG16	-	-	nrF	-			nrF											nrG12		
Najaguad149	rbG2	trmkG8	n/a	cG16	-	-	nrF	-			nrF											nrG12		
Najaguad115	rbG2	trmkG8	n/a	cG16	-	-	nrF	-			nrF											nrG12		
Najaguad109	rbG2	trmkG8	n/a	cG16	-	-	nrF	-			nrF											nrG12		nrG14
Najaguad110	n/a	trmkG8	n/a	cG16	-	-	nrF(p)	-			nrF											nrG12		nrG14
Najaguad116	rbG2	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes														nrG12		nrG14
Najaguad152	rbG2	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes														nrG12		
Najaguad153	rbG2	trmkG8	n/a	cG16	-	-	nrF(p)	-														nrG12		nrG14
Najaguad154	rbG2	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes														nrG12		nrG14
Najaguad156	rbG2	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes														nrG12		nrG14
Najaguad160	rbG2	trmkG8	n/a	cG16	-	-	nrF(p)	-														nrG12		nrG14
Najaguad163	rbG2	trmkG8	n/a	cG16	-	nrC	-	-														nrG12		
Najaguad164	rbG2	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes														nrG12		nrG14
Najaguad183	rbG2	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes														nrG12		
Najaguad184	rbG2	trmkG8	n/a	cG16	-	-	nrF(p)	-														nrG12		nrG14
Najaguad185	rbG2	trmkG8	n/a	cG16	-	-	nrF(p)	-														nrG12		nrG14
Najaguad192	rbG2	trmkG8	n/a	cG16	No	No	No	No														nrG12		
Najaguad195	rbG2	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes														nrG12		
Najaguad201	rbG2	trmkG8	n/a	cG16	No	No	No	No														nrG12		
Najaguad216	rbG2	trmkG8	n/a	cG16	No	No	No	No													nrG11	nrG12		
Najaguad003	n/a	trmkG8	n/a	cG16	n/a	n/a	n/a	n/a		nrC												nrG12		
Najaguad112	rbG2	trmkG8	trnG8	cG16	n/a	n/a	n/a	n/a	*	nrC												nrG12		

Najaguad217	rbG2	trmkG8	n/a	cG16	n/a	n/a	n/a	n/a		nrC													nrG12		
Najaguad221	rbG2	trmkG8	trnG8	cG16	-	nrC	-	-															nrG12		
Najaguad224	rbG2	trmkG8	n/a	cG16	No	No	No	No															nrG12		
Najaguad225	rbG2	trmkG8	n/a	cG16	No	No	No	No												nrG10			nrG12		
Najaguad226	rbG2	trmkG8	trnG8	cG16	nrUn	-	-	-												nrG10			nrG12		
Najaguad233	rbG2	trmkG8	n/a	cG16	Yes	Yes	Yes	Yes															nrG12		nrG14
Najaguad234	rbG2	trmkG8	n/a	cG16	-	-	nrF(p)	-															nrG12		nrG14
Najaguad243	rbG2	trmkG8	n/a	cG16	Faint	Faint	Faint	Faint															nrG12		
Najaguad256 (was Naja	rbG2	trmkG8	n/a	cG16	-	-	nrF(p)	-															nrG12		nrG14
Najaguad048	rbG2	trmkG8	trnG8	cG16	-	-	nrF(p)	-															nrG12		nrG14
Najaguad006	rbG2	trmkG7	n/a	cG15	n/a	n/a	n/a	n/a															nrG12		
Najaguad013	rbG2	trmkG7	trnG7	cG15	-	nrC	nrF	-	*					nrG3	nrG4										
Najaguad037	rbG2	trmkG7	n/a	cG15	-	nrC	-	-																nrG13	
Najaguad039	rbG2	trmkG7	n/a	cG15	No	No	No	No															nrG12	nrG13	
Najaguad051	rbG2	trmkG7	trnG7	cG15	No	No	No	No															nrG12	nrG13	
Najaguad142	rbG2	trmkG7	n/a	cG15	-	nrC	-	-															nrG12	nrG13	
Najaguad186	rbG2	trmkG7	n/a	cG15	No	No	No	No															nrG12	nrG13	
Najaguad194	rbG2	trmkG7	n/a	cG15	-	nrC	-	-															nrG12		
Najaguad199	rbG2	trmkG7	n/a	cG15	No	No	No	No															nrG12		
Najaguad206	rbG2	trmkG7	n/a	cG15	-	nrC	-	-															nrG12	nrG13	
Najaguad248	rbG2	trmkG7	n/a	cG15	-	nrC	-	-																nrG13	
Najaguad251	rbG2	trmkG7	n/a	cG15	No	No	No	No															nrG12		
Najaguad050	n/a	trmkG7	n/a	cG15	No	No	No	No															nrG12	nrG13	
Najaguad122	n/a	trmkG7	n/a	cG15	Yes	Yes	Yes	Yes															nrG12	nrG13	
Najaguad137	n/a	trmkG7	n/a	cG15	Yes	Yes	Yes	Yes															nrG12	nrG13	
Najaguad002	rbG2	trmkG7	n/a	cG15	n/a	n/a	n/a	-	*	nrC													nrG12		
Najaguad004	rbG2	trmkG7	n/a	cG15	n/a	n/a	n/a	-	*	nrC													nrG12		
Najaguad045	rbG2	trmkG7	n/a	cG15	-	nrC	nrF	-		nrC													nrG12		
Najaguad161/162	rbG1	trmkG6	trnG6a	cG14a	-	nrC	-	-															nrG12		
Najaguad247	rbG1	trmkG6	n/a	cG14	n/a	n/a	n/a	n/a															nrG12		
Najaguad219	rbG1	trmkG6	trnG6a	cG14a	-	nrC	-	-															nrG12		
Najaguad220	rbG1	trmkG6	n/a	cG14	-	nrC	-	-															nrG12		

Najaoliv098	rbG1	trmkG6	trnG6a	cG14a	-	-	nrF	-																	nrG14
Najaoliv227	rbG1	trmkG6	trnG6b	cG14b	-	-	nrF	-																	nrG14
Najaoliv238	rbG1	trmkG6	n/a	cG14	-	-	nrF	-																	nrG14
Najaoliv241	rbG1	trmkG6	trnG6a	cG14a	n/a	n/a	n/a	n/a																	nrG14
Najaoliv253	rbG1	trmkG6	n/a	cG14	-	-	nrF	-																	nrG14
Najaguad023	n/a	trmkG5	n/a	cG13	nrUn	-	-	-													nrG10				
Najaguad001	rbG1	trmkG5	trnG5	cG13	No	No	No	No	*									nrG7			nrG10				
Najaguad169	rbG1	trmkG5	n/a	cG13	n/a	n/a	n/a	n/a							nrG4						nrG10				
Najaguad212	rbG1	trmkG5	n/a	cG13	n/a	n/a	n/a	n/a						nrG3											
Najaguad214	rbG1	trmkG5	trnG5	cG13	nrUn	-	-	-						nrG3											
Najaguad031	rbG1	trmkG2	n/a	cG12	-	nrC	-	-												nrG9	nrG10				
Najaguad071	rbG1	trmkG2	n/a	cG12	No	No	No	No											nrG8						
Najaguad093	rbG1	trmkG2	n/a	cG12	No	No	No	No														nrG11			
Najaguad135	rbG1	trmkG2	trnG2	cG12	nrUn	-	-	-						nrG3							nrG10				
Najaguad170	rbG1	trmkG2	n/a	cG12	-	nrC	-	-										nrG7				nrG11			
Najaguad175	rbG1	trmkG2	n/a	cG12	nrUn	-	-	-												nrG9		nrG11			
Najaguad176	rbG1	trmkG2	trnG2	cG12	nrUn	-	-	-												nrG9		nrG11			
Najaguad178	rbG1	trmkG2	n/a	cG12	Yes	Yes	Yes	Yes						nrG3											
Najaguad063	rbG1	trmkG3	trnG3	cG11	nrUn	-	-	-											nrG8						
Najaguad072	rbG1	trmkG3	n/a	cG11	No	No	No	No												nrG9	nrG10				
Najaguad021	n/a	trmkG1	n/a	cG10	No	No	No	No										nrG7							
Najaguad016	n/a	trmkG1	n/a	cG10	n/a	n/a	n/a	n/a													nrG10				
Najaguad089	n/a	trmkG1	n/a	cG10	No	No	No	No												nrG9	nrG10				
Najaguad012	rbG1	trmkG1	trnG1	cG10	Faint	Faint	Faint	Faint	*									nrG7				nrG11			
Najaguad017	rbG1	trmkG1	n/a	cG10	-	-	-	nrG1				nrG1						nrG7			nrG10				
Najaguad018	rbG1	trmkG1	trnG1	cG10	nrUn	-	-	-													nrG10				
Najaguad020	rbG1	trmkG1	n/a	cG10	nrUn	-	-	-										nrG7							
Najaguad022	rbG1	trmkG1	n/a	cG10	No	No	No	No													nrG10				
Najaguad032	rbG1	trmkG1	n/a	cG10	-	nrC	nrF	-													nrG10				
Najaguad043	rbG1	trmkG1	trnG1	cG10	nrUn	-	-	-										nrG7			nrG10				

Najaguad044	rbG1	trmkG1	n/a	cG10	No	No	No	No	*												nrG11	nrG12		
Najaguad054	rbG1	trmkG1	n/a	cG10	Yes	Yes	Yes	Yes											nrG9	nrG10				
Najaguad055	rbG1	trmkG1	n/a	cG10	Yes	Yes	Yes	Yes											nrG9	nrG10				
Najaguad056	rbG1	trmkG1	n/a	cG10	No	No	No	No										nrG8		nrG10				
Najaguad057	rbG1	trmkG1	n/a	cG10	No	No	No	No											nrG9	nrG10				
Najaguad058	rbG1	trmkG1	n/a	cG10	No	No	No	No						nrG3				nrG8						
Najaguad059	rbG1	trmkG1	n/a	cG10	No	No	No	No							nrG4		nrG7							
Najaguad060	rbG1	trmkG1	n/a	cG10	n/a	n/a	n/a	n/a							nrG4					nrG10				
Najaguad061	rbG1	trmkG1	n/a	cG10	nrUn	-	-	-													nrG11			
Najaguad064	rbG1	trmkG1	n/a	cG10	n/a	n/a	n/a	n/a											nrG9		nrG11			
Najaguad066	rbG1	trmkG1	n/a	cG10	No	No	No	No											nrG9	nrG10				
Najaguad067	rbG1	trmkG1	n/a	cG10	nrUn	-	-	-											nrG9	nrG10				
Najaguad068	rbG1	trmkG1	n/a	cG10	No	No	No	No									nrG7			nrG10				
Najaguad069	rbG1	trmkG1	n/a	cG10	Faint	Faint	Faint	Faint									nrG7			nrG10				
Najaguad070	rbG1	trmkG1	n/a	cG10	No	No	No	No										nrG8		nrG10				
Najaguad073	rbG1	trmkG1	n/a	cG10	No	No	No	No											nrG9	nrG10				
Najaguad076	rbG1	trmkG1	trnG1	cG10	No	No	No	No													nrG11			
Najaguad077	rbG1	trmkG1	n/a	cG10	No	No	No	No										nrG8		nrG10				
Najaguad078	rbG1	trmkG1	n/a	cG10	No	No	No	No										nrG8		nrG10				
Najaguad079	rbG1	trmkG1	n/a	cG10	No	No	No	No										nrG8		nrG10				
Najaguad080	rbG1	trmkG1	n/a	cG10	nrUn	-	-	-										nrG8		nrG10				
Najaguad082	rbG1	trmkG1	n/a	cG10	nrUn	-	-	-										nrG8		nrG10				
Najaguad083	rbG1	trmkG1	n/a	cG10	Yes	Yes	Yes	Yes											nrG9	nrG10				
Najaguad084	rbG1	trmkG1	n/a	cG10	Faint	Faint	Faint	Faint										nrG8		nrG10				
Najaguad085	rbG1	trmkG1	n/a	cG10	Faint	Faint	Faint	Faint										nrG8		nrG10				
Najaguad086	rbG1	trmkG1	n/a	cG10	Yes	Yes	Yes	Yes											nrG9	nrG10				
Najaguad087	rbG1	trmkG1	n/a	cG10	No	No	No	No	*									nrG8	nrG9					
Najaguad088	rbG1	trmkG1	n/a	cG10	No	No	No	No										nrG8		nrG10				
Najaguad090	rbG1	trmkG1	n/a	cG10	No	No	No	No												nrG10				
Najaguad091	rbG1	trmkG1	n/a	cG10	No	No	No	No											nrG9	nrG10				
Najaguad092	rbG1	trmkG1	n/a	cG10	No	No	No	No										nrG8		nrG10				
Najaguad096	rbG1	trmkG1	n/a	cG10	No	No	No	No											nrG9	nrG10				
Najaguad148	rbG1	trmkG1	n/a	cG10	No	No	No	No										nrG8						
Najaguad150	rbG1	trmkG1	n/a	cG10	-	nrC	-	-											nrG9	nrG10				

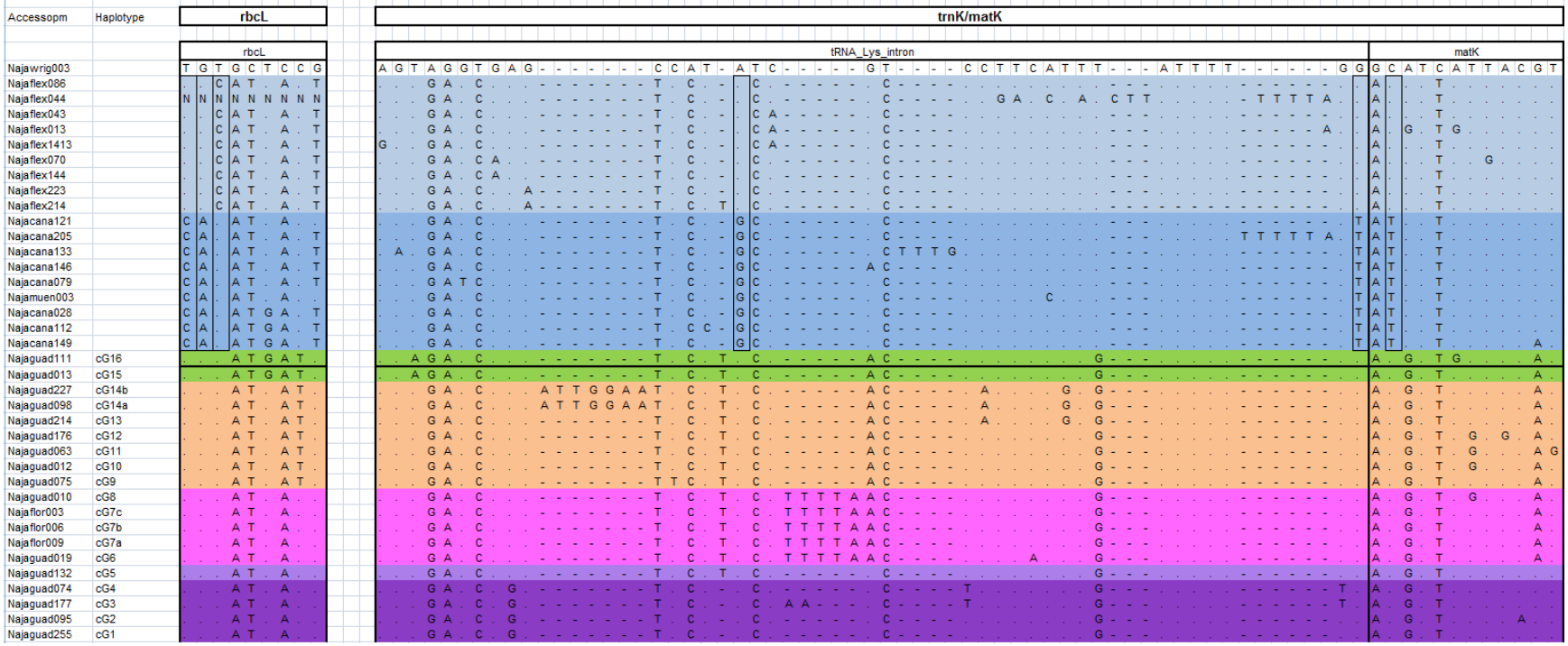
Najaguad168	rbG1	trmkG1	n/a	cG10	Faint	Faint	Faint	Faint							nrG4						nrG10				
Najaguad171	rbG1	trmkG1	n/a	cG10	-	nrC	nrF	-										nrG7				nrG11			
Najaguad172	rbG1	trmkG1	n/a	cG10	-	nrC	-	-														nrG11			
Najaguad173	rbG1	trmkG1	n/a	cG10	Yes	Yes	Yes	Yes														nrG11			
Najaguad174	rbG1	trmkG1	n/a	cG10	Faint	Faint	Faint	Faint												nrG9		nrG11			
Najaguad223	rbG1	trmkG1	trnG1	cG10	nrUn	-	-	-													nrG10				
Najaguad228	rbG1	trmkG1	trnG1	cG10	nrUn	-	-	-										nrG7			nrG10				
Najaguad229	rbG1	trmkG1	n/a	cG10	No	No	No	No												nrG9		nrG11			
Najaguad235	rbG1	trmkG1	n/a	cG10	-	-	-	nrG1b				nrG1													
Najaguad236	rbG1	trmkG1	n/a	cG10	No	No	No	No						nrG3								nrG11			
Najaguad237	rbG1	trmkG1	n/a	cG10	No	No	No	No										nrG7			nrG10				
Najaguad245	rbG1	trmkG1	trnG1	cG10	No	No	No	No											nrG8		nrG10				
Najaguad008	rbG1	trmkG1	trnG1	cG10	No	No	No	No													nrG10		nrG12		
Najaguad145	rbG1	trmkG1	n/a	cG10	n/a	n/a	n/a	n/a												nrG9	nrG10				
Najaguad033	n/a	trmkG1	n/a	cG10	No	No	No	No												nrG9	nrG10				
Najaguad075	rbG1	trmkG4	trnG4	cG9	No	No	No	No														nrG11			
Najaguad010	rbFL1	trmkFL3	trnFL3	cG8	-	-	-	nrG1				nrG1										nrG11			
Najaflor002	n/a	trmkFL1	n/a	cG7	-	-	-	nrG1	*			nrG1			nrG4										
Najaflor003	rbFL1	trmkFL1	trnFL1b	cG7c	-	-	-	nrG1	*			nrG1			nrG4										
Najaflor006	rbFL1	trmkFL1	trnFL1c	cG7b	-	-	-	nrG1				nrG1										nrG11			
Najaflor007	rbFL1	trmkFL1	n/a	cG7	-	-	-	nrG1				nrG1													
Najaflor009	rbFL1	trmkFL1	trnFL1a	cG7a	-	-	-	nrG1				nrG1			nrG4							nrG11			
Najaflor010	rbFL1	trmkFL1	n/a	cG7	-	-	-	nrG1				nrG1		nrG3								nrG11			
Najaflor014	rbFL1	trmkFL1	n/a	cG7	-	-	-	nrG1				nrG1		nrG3								nrG11			
Najaflor015 (was Najag	rbFL1	trmkFL1	n/a	cG7	-	-	-	nrG1				nrG1													
Najaflor016 (was Najag	rbFL1	trmkFL1	n/a	cG7	-	-	-	nrG1				nrG1						nrG7				nrG11			
Najaguad019	rbFL1	trmkFL2	trnFL2	cG6	-	-	-	nrG1b	*			nrG1						nrG7							
Najaguad132	rbFL1	trmkFL8	trnFL8	cG5	-	nrC	nrF	-					nrG2									nrG11			
Najaguad133/134	rbFL1	trmkFL8	trnFL8	cG5	-	nrC	nrF	-					nrG2									nrG11			

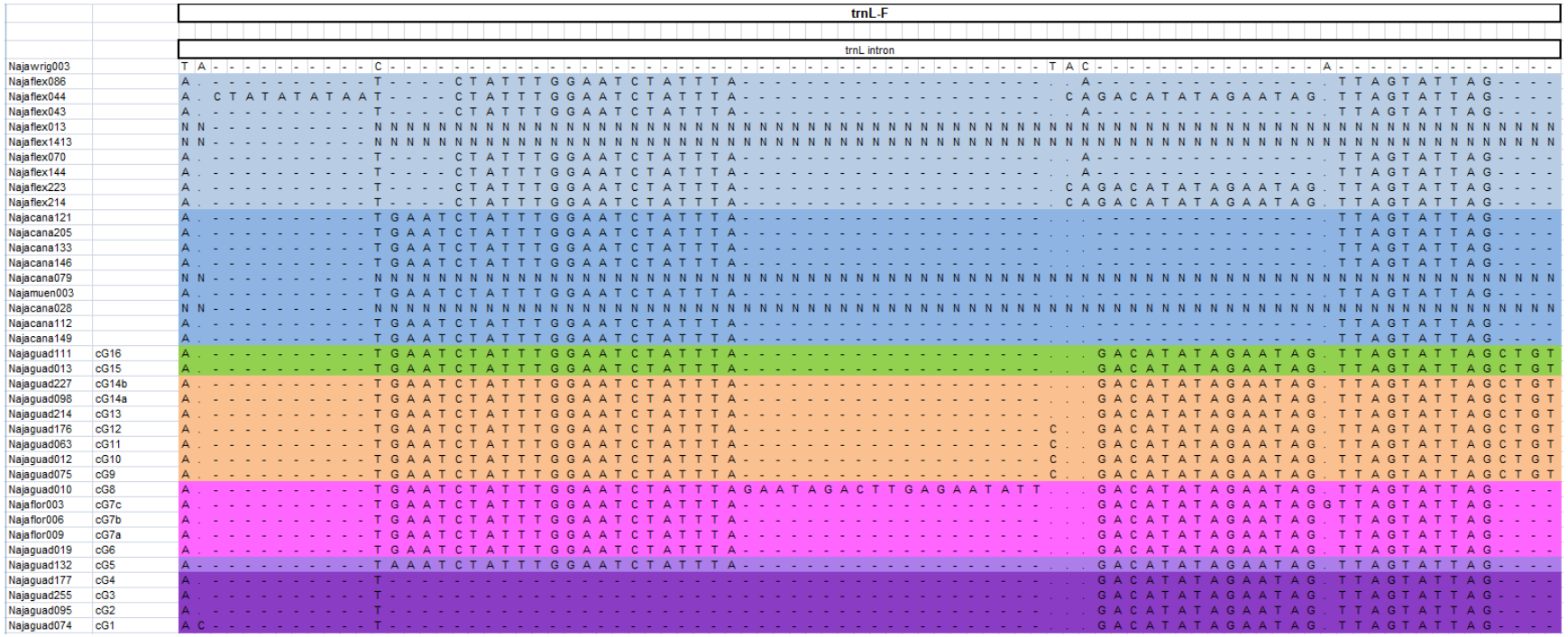
Najaguad074	rbFL1	trmkFL5	trnFL5	cG4	No	No	No	No												nrG9	nrG10				
Najaguad062	rbFL1	trmkFL4	trnFL4	cG3	-	nrC	-	-								nrG5									
Najaguad065	rbFL1	trmkFL4	trnFL4	cG3	nrUn	-	-	-											nrG8	nrG9					
Najaguad124	rbFL1	trmkFL4	trnFL4	cG3	-	nrC	nrF	-												nrG9					
Najaguad177	rbFL1	trmkFL4	trnFL4	cG3	nrUn	-	-	-												nrG9					
Najaguad094	rbFL1	trmkFL6	trnFL6	cG2	-	nrC	-	-												nrG9					
Najaguad095	rbFL1	trmkFL6	trnFL6	cG2	No	No	No	No												nrG9					
Najaguad196/197	rbFL1	trmkFL6	trnFL6	cG2	No	No	No	No												nrG9					
Najaguad197	rbFL1	trmkFL6	trnFL6	cG2	No	No	No	No												nrG9					
Najaguad244	rbFL1	trmkFL6	trnFL6	cG2	No	No	No	No												nrG9					
Najaguad255 (was Naja	rbFL1	trmkFL7	trnFL7	cG1	No	No	No	No									nrG6								
Najaflor004	n/a	n/a	n/a		n/a	n/a	n/a	n/a				nrG1													
Najaflor005	n/a	n/a	n/a		-	-	-	nrG1				nrG1													
Najaflor008	n/a	n/a	n/a		n/a	n/a	n/a	n/a				nrG1													
Najaguad011	n/a	n/a	n/a		n/a	n/a	n/a	n/a																nrG12		nrG14
Najaguad036	n/a	n/a	n/a		n/a	n/a	n/a	n/a																nrG12		
Najaguad126	n/a	n/a	n/a		Faint	Faint	Faint	Faint																nrG12		
Najaguad151	n/a	n/a	n/a		Yes	Yes	Yes	Yes																nrG12	nrG13	
Najaguad159	n/a	n/a	n/a		Yes	Yes	Yes	Yes																nrG12		nrG14
Najaguad182	n/a	n/a	n/a		Yes	Yes	Yes	Yes																nrG12		nrG14
Najaguad187	n/a	n/a	n/a		Yes	Yes	Yes	Yes																nrG12	nrG13	
Najaguad189	n/a	n/a	n/a		Yes	Yes	Yes	Yes																nrG12	nrG13	
Najaguad190	n/a	n/a	n/a		Yes	Yes	Yes	Yes																nrG12	nrG13	
Najaguad191	n/a	n/a	n/a		Yes	Yes	Yes	Yes																nrG12		
Najaguad204	n/a	n/a	n/a		n/a	n/a	n/a	n/a																nrG12		
Najaguad205	n/a	n/a	n/a		n/a	n/a	n/a	n/a																nrG12		
Najaguad242	n/a	n/a	n/a		No	No	No	No																nrG12		
Najaguad180	rbG2	n/a	n/a		No	No	No	No																nrG12		
Najaoliv232	n/a	n/a	n/a		n/a	n/a	n/a	n/a																		nrG14

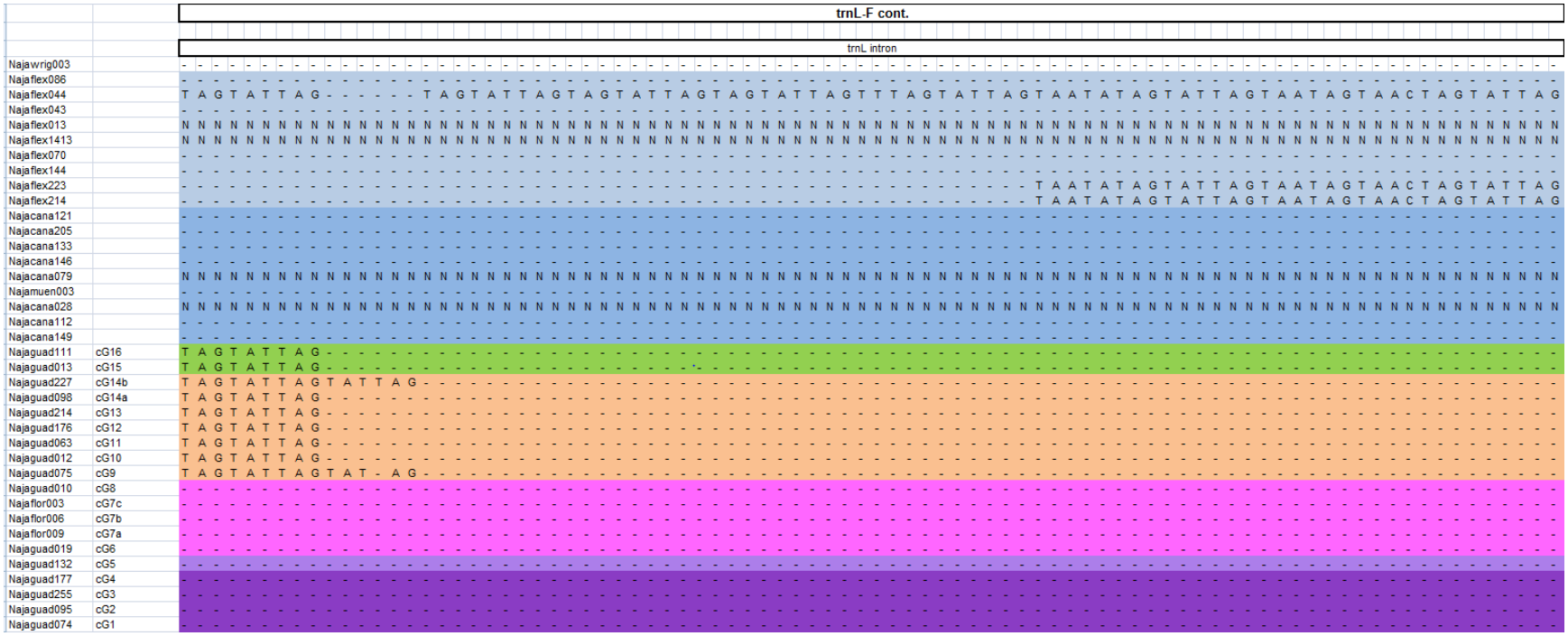
Najaoliv240	n/a	n/a	n/a		-	-	nrF	-																	nrG14
Najaoliv230	n/a	n/a	n/a		-	-	nrF(p)	-																	nrG14
Najaguad146	n/a	n/a	n/a		No	No	No	No									nrG6								
Najaguad081	n/a	n/a	n/a		No	No	No	No											nrG8		nrG10				
Najaguad024	n/a	n/a	n/a		n/a	n/a	n/a	n/a													nrG10				
Najaguad147	n/a	n/a	n/a		No	No	No	No														nrG11			
Najaguad198	n/a	n/a	n/a		No	No	No	No												nrG9					
Najaguad202	n/a	n/a	n/a		No	No	No	No												nrG9	nrG10				
Najaguad207	n/a	n/a	n/a		No	No	No	No													nrG10	nrG11			
Najaguad208	n/a	n/a	n/a		No	No	No	No													nrG10				
Najaguad209	n/a	n/a	n/a		No	No	No	No												nrG9					
Najaguad210	n/a	n/a	n/a		No	No	No	No													nrG10	nrG11			
Najaguad211	n/a	n/a	n/a		No	No	No	No												nrG9					
Najaguad213	n/a	n/a	n/a		Faint	Faint	Faint	Faint												nrG9		nrG11			
Najaguad222	n/a	n/a	n/a		No	No	No	No													nrG10	nrG11			
Najaguad028	rbFL1	n/a	n/a		No	No	No	No												nrG9					
Najaguad034	rbG1	n/a	n/a		No	No	No	No													nrG10				
Najaguad040	n/a	trmKC1	n/a		n/a	n/a	n/a	n/a	*	nrC													nrG12	nrG13	
Najaguad041	rbC1	trmKC1	n/a		-	nrC	-	-	*		nrF												nrG12		
Najaguad046	rbC1	trmKC1	n/a		-	nrC	-	-															nrG12		
Najaguad188	rbC1	trmKC1	n/a		n/a	n/a	n/a	n/a															nrG12		

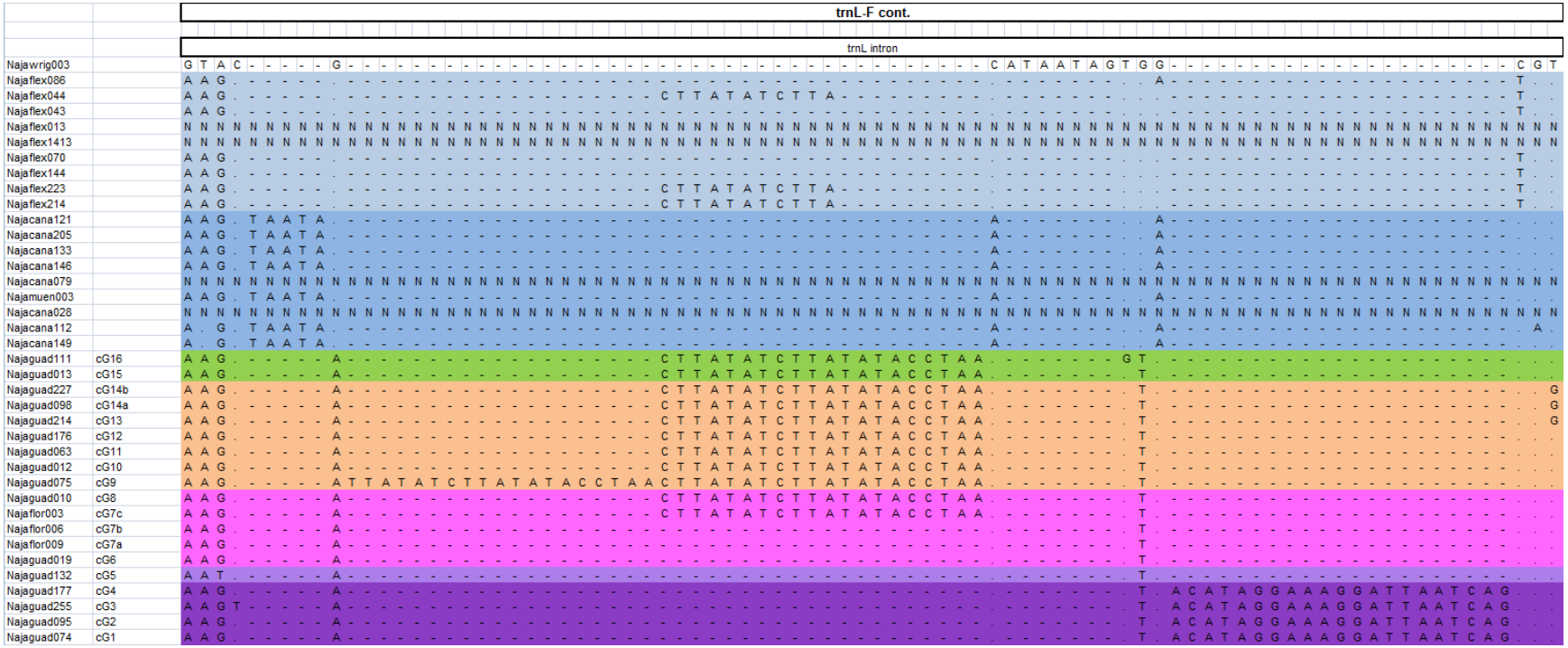
APPENDIX D

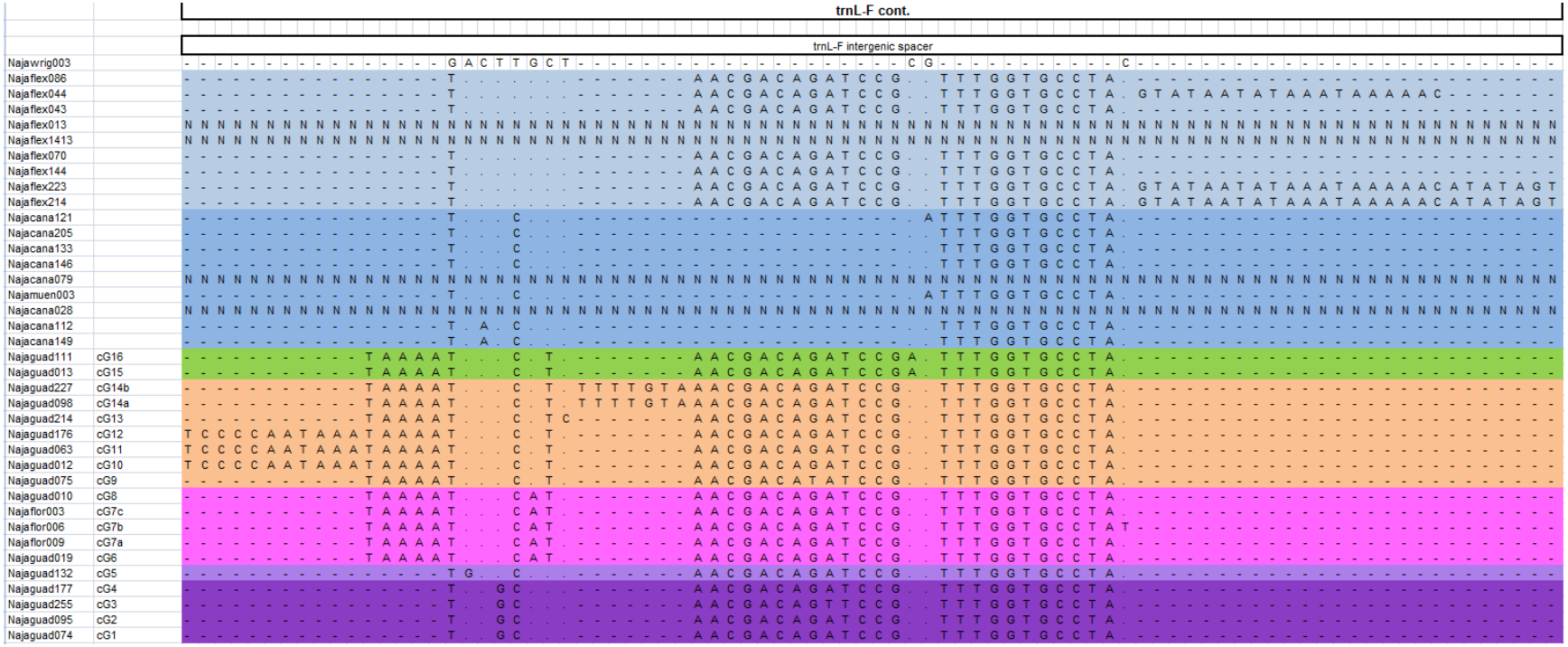
Sequence variation for the chloroplast haplotypes obtained with the rbcL, trnK/matK and trnL-F regions
Six boxed regions indicate conserved differences between N. flexilis and N. canadensis (Les et al. 2015)

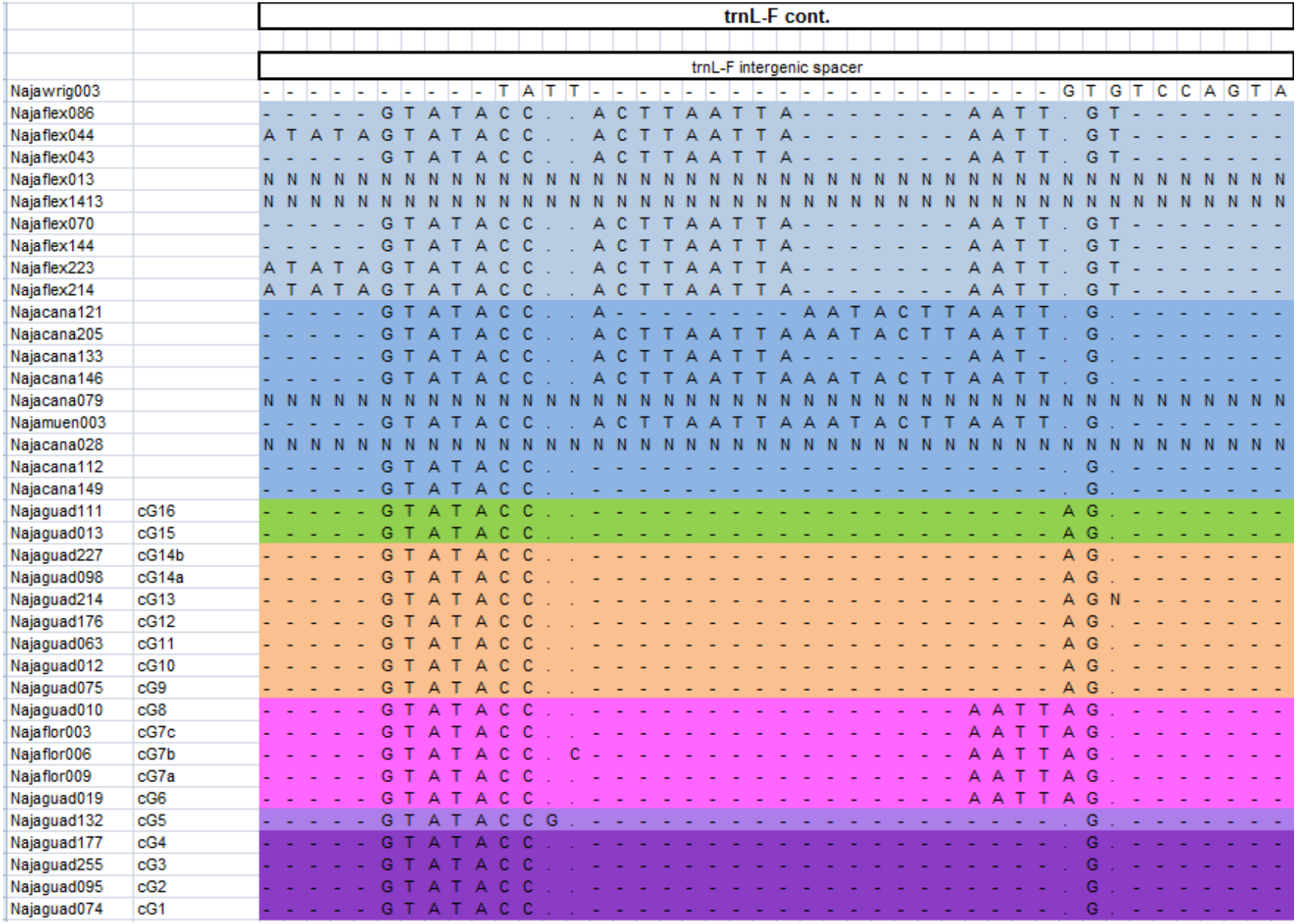



trnL-F
trnL intron

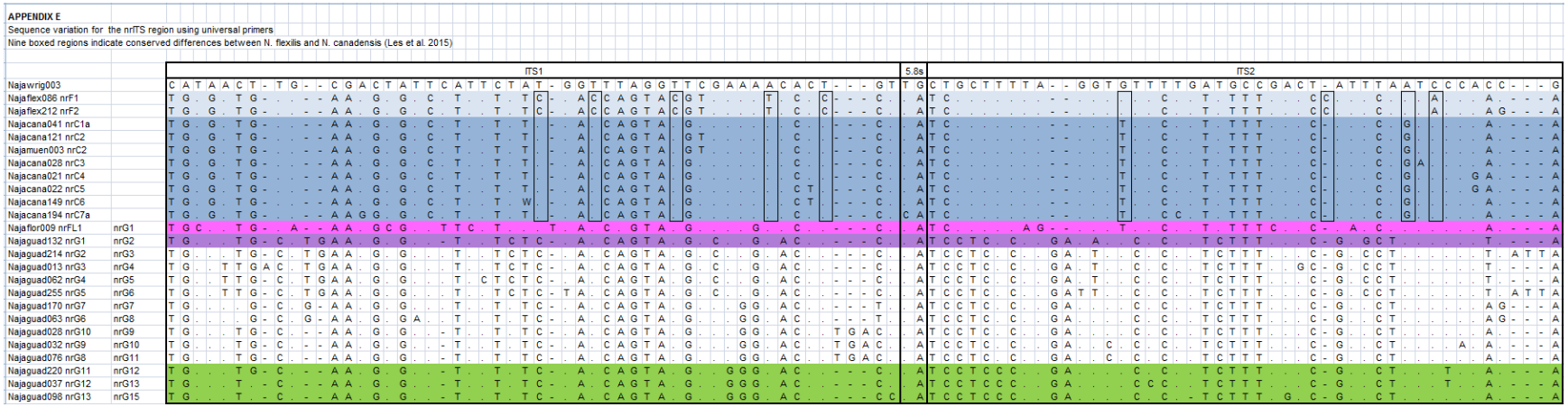

trnL-F cont.
trnL intron
Najawrig003
Najaflex086
Najaflex044
Najaflex043
Najaflex013
Najaflex1413
Najaflex070
Najaflex144
Najaflex223
Najaflex214
Najacana121
Najacana205
Najacana133
Najacana146
Najacana079
Najamuen003
Najacana028
Najacana112
Najacana149
Najaguad111 cG16
Najaguad013 cG15
Najaguad227 cG14b
Najaguad098 cG14a
Najaguad214 cG13
Najaguad176 cG12
Najaguad063 cG11
Najaguad012 cG10
Najaguad075 cG9
Najaguad010 cG8
Najaflor003 cG7c
Najaflor006 cG7b
Najaflor009 cG7a
Najaguad019 cG6
Najaguad132 cG5
Najaguad177 cG4
Najaguad255 cG3
Najaguad095 cG2
Najaguad074 cG1


trnL-F cont.
trnL intron


trnL-F cont.
trnL-F intergenic spacer


trnL-F cont.
trnL-F intergenic spacer
Najawrig003
Najaflex086
Najaflex044
Najaflex043
Najaflex013
Najaflex1413
Najaflex070
Najaflex144
Najaflex223
Najaflex214
Najacana121
Najacana205
Najacana133
Najacana146
Najacana079
Najamuen003
Najacana028
Najacana112
Najacana149
Najaguad111 cG16
Najaguad013 cG15
Najaguad227 cG14b
Najaguad098 cG14a
Najaguad214 cG13
Najaguad176 cG12
Najaguad063 cG11
Najaguad012 cG10
Najaguad075 cG9
Najaguad010 cG8
Najaflor003 cG7c
Najaflor006 cG7b
Najaflor009 cG7a
Najaguad019 cG6
Najaguad132 cG5
Najaguad177 cG4
Najaguad255 cG3
Najaguad095 cG2
Najaguad074 cG1

Appendix E


APPENDIX E
Sequence variation for the nrITS region using universal primers
Nine boxed regions indicate conserved differences between N. flexilis and N. canadensis (Les et al. 2015)
ITS1
5.8s
ITS2
Najawrig003
Najaflex086 nrF1
Najaflex212 nrF2
Najacana041 nrC1a
Najacana121 nrC2
Najamuen003 nrC2
Najacana028 nrC3
Najacana021 nrC4
Najacana022 nrC5
Najacana149 nrC6
Najacana194 nrC7a
Najaflor009 nrFL1 nrG1
Najaguad132 nrG1 nrG2
Najaguad214 nrG2 nrG3
Najaguad013 nrG3 nrG4
Najaguad062 nrG4 nrG5
Najaguad255 nrG5 nrG6
Najaguad170 nrG7 nrG7
Najaguad063 nrG6 nrG8
Najaguad028 nrG10 nrG9
Najaguad032 nrG9 nrG10
Najaguad076 nrG8 nrG11
Najaguad220 nrG11 nrG12
Najaguad037 nrG12 nrG13
Najaguad098 nrG13 nrG15

APPENDIX F

APPENDIX F.

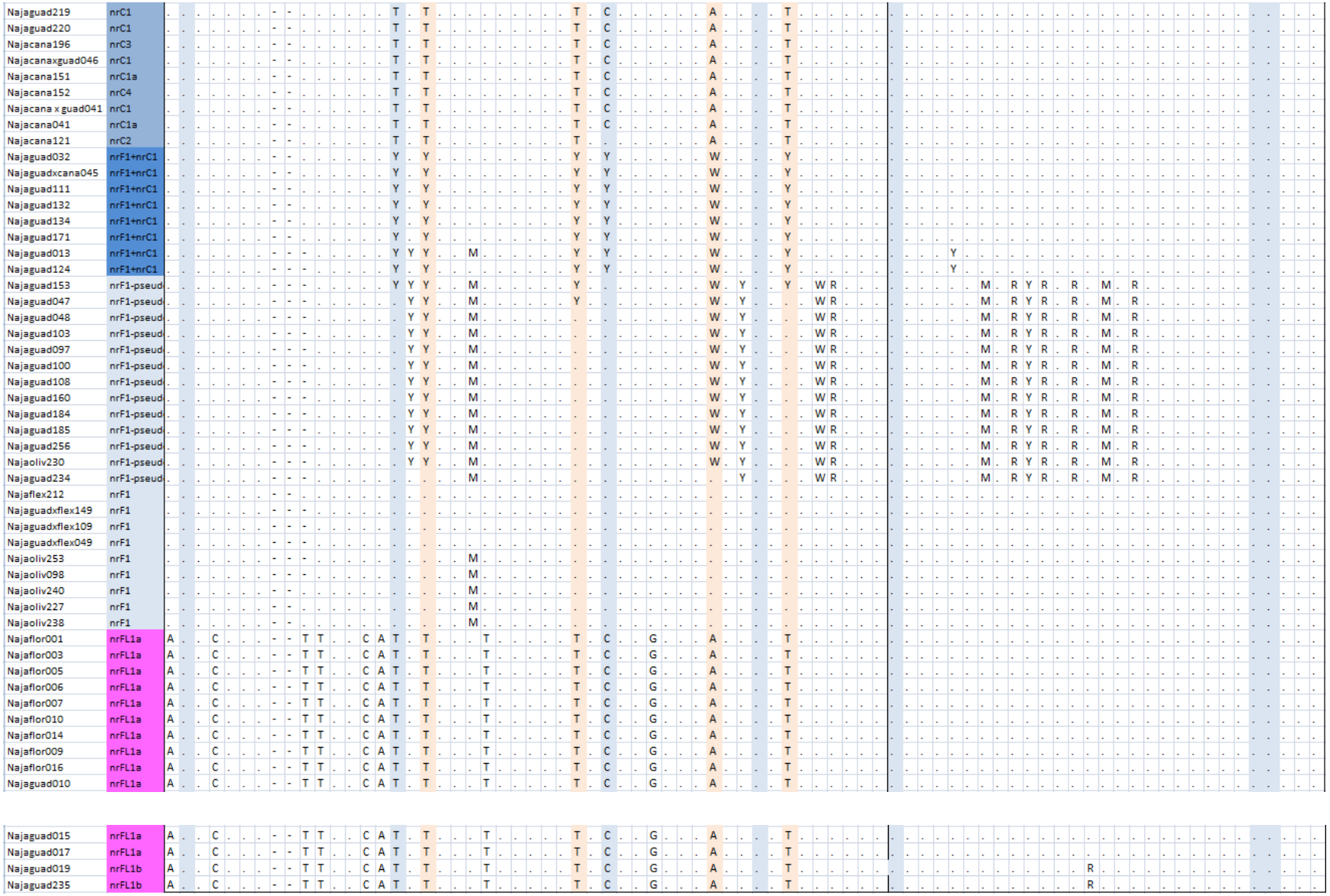
Table showing substitutions found with the Nflex-ITS repeat-specific primers - relative to N. flexilis (Splitstree analysis)

(orange) Substitutions that separate N. flexilis and N. canadensis (Les et al. 2015)

(light blue) Substitutions previously found within the N. canadensis (nrC1-nrC7) clade (Les et al. 2015)

		ITS1																																													
Najaflex086		G	A	A	T	C	G	T	-	-	C	C	G	C	T	T	C	C	C	G	C	C	A	G	T	G	A	A	C	G	T	G	C	A	C	G	T	T	G	C	C	A	C	C	T	G	C	G	C
Najaguad135	nrUn	N	.	R	.	.	M	K	-	-	-	.	R	.	.	.	T	.	T	.	.	.	.	R	Y	.	.	W	T	.	C	.	Y	.	Y	R	Y	A	.	.	.	R	T	S	.	.	.	.	.
Najaguad214	nrUn	N	.	R	.	.	M	K	-	-	-	.	R	.	.	.	T	.	T	.	.	.	.	R	Y	.	.	W	T	.	C	.	Y	.	Y	R	Y	A	.	.	.	R	T	S	.	.	.	.	.
Najaguad175	nrUn	.	.	.	.	.	.	.	-	-	.	.	R	.	.	.	T	.	T	.	M	.	.	.	Y	.	R	.	T	S	Y	.	Y	.	Y	.	Y	A	.	.	.	R	T	S	.	.	.	.	.
Najaguad177	nrUn	.	.	.	.	.	.	.	-	-	.	.	R	.	.	.	T	.	T	.	M	.	.	.	Y	.	R	.	T	S	Y	.	Y	.	Y	.	Y	A	.	.	.	R	T	S	.	.	.	.	.
Najaguad176	nrUn	.	.	.	.	.	.	.	-	-	.	.	R	.	.	.	T	.	T	.	M	.	.	.	Y	.	R	.	T	S	Y	.	Y	.	Y	.	Y	A	.	.	.	R	T	S	.	.	.	.	.
Najaguad065	nrUn	N	.	.	.	.	.	.	-	-	.	.	A	.	.	.	T	.	T	R	M	.	.	.	C	.	G	.	T	C	C	.	T	.	T	.	C	A	.	.	.	G	T	G	.	.	.	.	.
Najaguad228	nrUn	.	.	.	.	Y	.	.	C	C	.	.	A	.	.	.	T	.	T	.	.	.	.	.	C	R	.	.	T	.	C	R	T	.	T	R	C	A	R	.	.	G	T	G	.	.	.	R	.
Najaguad226	nrUn	.	.	.	.	Y	.	.	C	C	.	.	A	.	.	.	T	.	T	.	.	.	.	.	C	R	.	.	T	.	C	R	T	.	T	N	C	A	R	.	.	G	T	G	.	.	.	R	.
Najaguad063	nrUn	.	.	.	.	Y	.	.	-	-	.	.	A	.	.	.	T	.	T	.	.	.	.	.	C	.	.	.	T	.	Y	.	T	.	T	R	C	A	.	.	.	G	T	G	.	R	Y	R	Y
Najaguad020	nrUn	.	.	.	.	T	.	.	-	C	.	.	A	.	.	.	T	.	T	.	.	.	.	.	C	.	.	.	T	.	C	.	T	.	T	A	C	A	.	.	.	G	T	G	.	.	.	.	.
Najaguad080	nrUn	.	.	.	.	Y	.	.	-	C	.	.	A	.	.	.	T	.	T	.	.	.	.	.	C	.	.	.	T	.	C	.	T	.	T	R	C	A	.	.	.	G	T	G	.	.	.	R	.
Najaguad082	nrUn	.	.	.	.	Y	.	.	-	C	.	.	A	.	.	.	T	.	T	.	.	.	.	.	C	.	.	.	T	.	C	R	T	.	T	R	C	A	.	.	.	G	T	G	.	.	.	R	.
Najaguad061	nrUn	.	.	.	.	Y	.	.	-	C	.	.	A	.	.	.	T	.	T	.	.	.	.	.	C	.	.	.	T	.	C	R	T	.	T	R	C	A	.	.	.	G	T	G	.	.	.	R	.
Najaguad223	nrUn	.	.	.	.	Y	.	.	-	C	.	.	A	.	.	.	T	.	T	.	.	.	.	.	C	.	.	.	T	.	C	R	T	.	T	R	C	A	.	.	.	G	T	G	.	.	.	R	.
Najaguad018	nrUn	.	.	.	.	Y	.	.	-	C	.	.	A	.	.	.	T	.	T	.	.	.	.	.	C	.	.	.	T	.	C	R	T	.	T	R	C	A	.	.	.	G	T	G	.	.	.	R	.
Najaguad023	nrUn	.	.	.	.	Y	.	.	-	C	.	.	A	.	.	.	T	.	T	.	.	.	.	.	C	.	.	.	T	.	C	R	T	.	T	R	C	A	.	.	.	G	T	G	.	.	.	R	.
Najaguad043	nrUn	.	.	.	.	T	.	.	-	-	.	.	A	T	.	.	T	.	T	.	.	.	.	.	C	.	.	.	T	.	C	.	T	.	T	A	C	A	.	.	.	G	T	G	.	.	.	.	.
Najaguad067	nrUn	.	.	.	.	T	.	.	-	-	.	.	A	T	.	.	T	.	T	.	.	.	.	.	C	.	.	.	T	.	C	.	T	.	T	A	C	A	.	.	.	G	T	G	.	.	.	.	.
NcanaIRE_NAH082	nrC7	.	G	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
NcanaIRE_POL139	nrC7	.	G	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
NcanaSCO_RUN380	nrC7	.	G	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najacana225	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad172	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najamuen003	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najacana248	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad162	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	Y	.	.	.	.	.	.	.	.	.	T	.	Y	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad161	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	Y	.	.	.	.	.	.	.	.	.	T	.	Y	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad163	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	Y	.	.	.	.	.	.	.	.	.	T	.	Y	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad194	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	Y	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad114	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	Y	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad062	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	Y	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad170	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	Y	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad031	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	Y	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najacana149	nrC6	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	T	.	T	.	.	.	.	.	.
Najacana119	nrC5	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	T	.	T	.	.	.	.	.	.
Najacana235	nrC5	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	T	.	T	.	.	.	.	.	.
Najaguad150	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	T	.	T	.	.	.	.	.	.
Najacana1417	nrC4	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad037	nrC1	.	.	.	.	.	.	.	-	-	-	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad206	nrC1	N	N	N	N	N	N	N	-	-	-	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad142	nrC1	N	.	.	.	.	.	.	-	-	-	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad094	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.
Najaguad221	nrC1	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	T	.	T	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.

		5.8s																												
Najaflex086		T	C	C	G	C	T	C	C	G	C	G	A	G	G	C	C	G	T	C	G	G	C	G	G	A	G	C	C	T
Najaguad135	nrUn	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.	.	.	Y	.	R	R	Y	.	R	.	.	.	.	.
Najaguad214	nrUn	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.	.	.	Y	.	R	R	Y	.	R	.	.	.	.	.
Najaguad175	nrUn	.	Y	.	.	.	.	.	.	.	.	.	W	.	.	.	.	.	.	.	.	.	.	K	.	.	.	.	Y	W
Najaguad177	nrUn	.	Y	.	.	.	.	.	.	.	.	.	W	.	.	.	.	.	.	.	.	.	.	K	.	.	.	.	Y	W
Najaguad176	nrUn	.	Y	.	.	.	.	.	.	.	.	.	W	.	.	.	.	.	.	.	.	.	.	K	.	.	.	.	Y	W
Najaguad065	nrUn	.	Y	.	.	.	.	.	.	.	.	.	W	.	.	.	.	.	.	.	.	.	.	K	.	.	.	.	Y	W
Najaguad228	nrUn	.	Y	.	.	.	W	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	K	.	.	.	.	Y	.
Najaguad226	nrUn	.	Y	.	.	.	W	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	N	.	.	.	.	Y	.
Najaguad063	nrUn	.	Y	.	R	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	Y	.	.	.	K	.	.	.	.	.	.
Najaguad020	nrUn	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	A	.	.
Najaguad080	nrUn	.	Y	.	.	.	.	.	.	.	.	.	.	.	K	.	Y	.	.	.	.	.	.	K	R	.	.	M	Y	.
Najaguad082	nrUn	.	Y	Y	R	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	K	.	.	.	M	Y	.
Najaguad061	nrUn	.	Y	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	K	.	.	.	M	Y	.
Najaguad223	nrUn	.	Y	Y	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	K	.	.	.	.	Y	.
Najaguad018	nrUn	.	Y	Y	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	K	.	.	.	.	Y	.
Najaguad023	nrUn	.	Y	Y	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	K	.	.	.	.	Y	.
Najaguad043	nrUn	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
Najaguad067	nrUn	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
NcanaIRE_NAH082	nrC7	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NcanaIRE_POL139	nrC7	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
NcanaSCO_RUN380	nrC7	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najacana225	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad172	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najamuen003	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najacana248	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad162	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad161	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad163	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad194	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad114	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	N	N	N	N
Najaguad062	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad170	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad031	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najacana149	nrC6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najacana119	nrC5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najacana235	nrC5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad150	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najacana1417	nrC4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	K	.	.	.
Najaguad037	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad206	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad142	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad094	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Najaguad221	nrC1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.



Appendix G

Seed images

Pericarp removed prior to photographing by rubbing seed on adhesive tape

R/S = repeat specific primers

Universal = universal nrITS primers

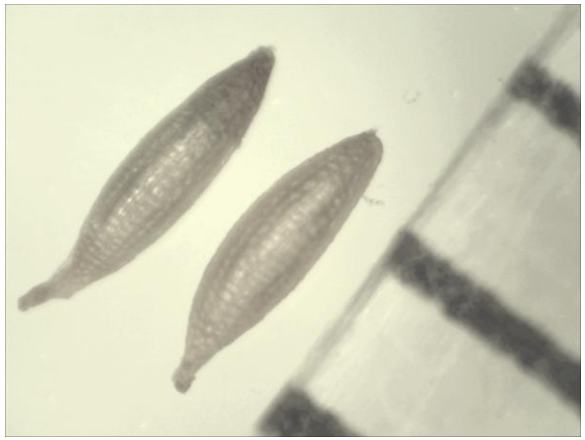
Najaguad096
Female flower & Male dehisced

cG1

Only 1 specimen
Najaguad255
Honduras
cG1
Universal: nrG6
R/S: amp guad not flex

Nil with flex primers
nrG5

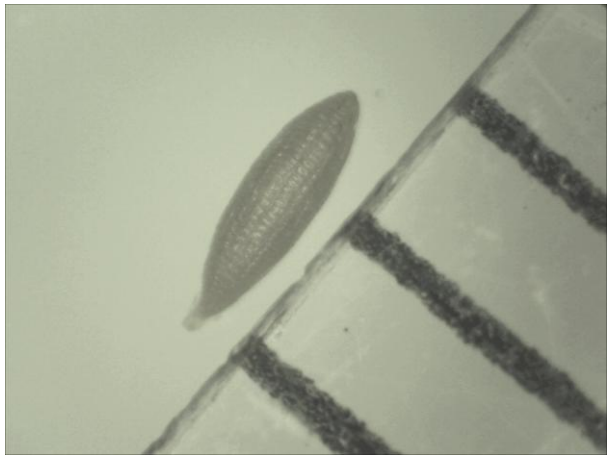
Najaguad146
Costa Rica
Universal: nrG6
No chloroplast loci
R/S: amp guad not flex
Lots seed and male flowers
Pericarp very thin



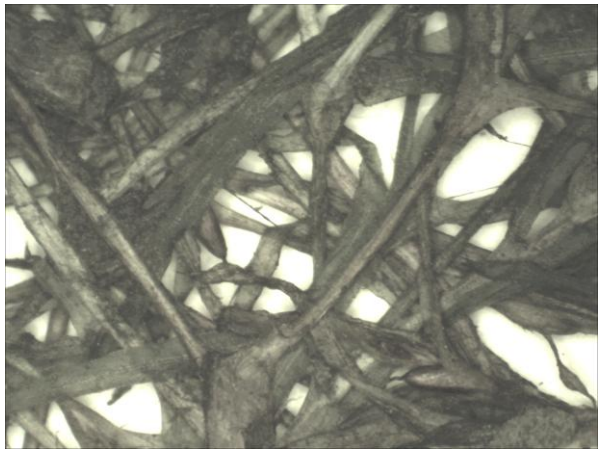
cG2 -4 out of 5 have seed

Najaguad094
Nebraska
cG2
Universal: nrG9
R/S: nrC
Lots seed

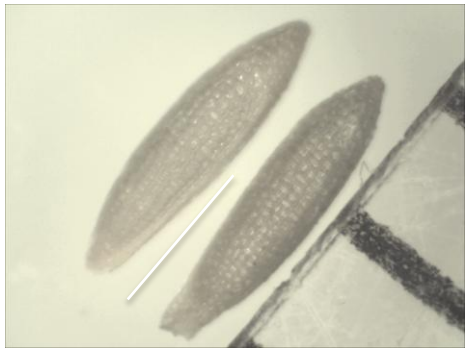
Najaguad244
South Dakota
Universal: nrG9
R/S : amp guad not flex
Lots seed

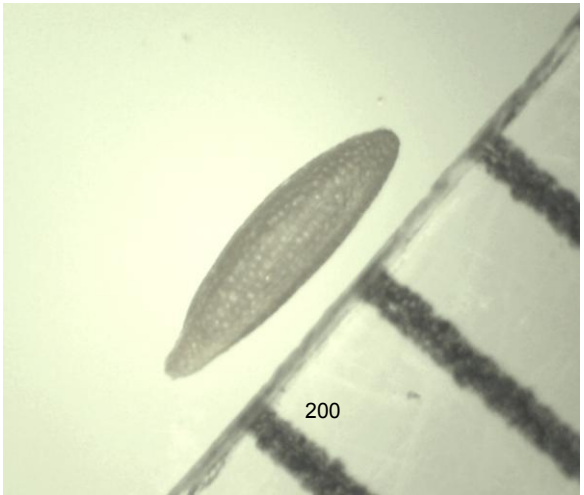


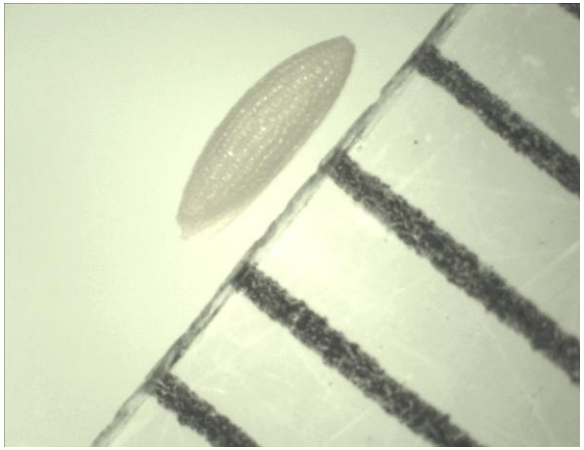
Najaguad196/197
South Dakota
Universal: nrG9
R/S: amp guad not flex
Lots seed



Najaguad028
Colorado
rbFL1 - no trmK
Universal: nrG9
R/S: amp guad not flex



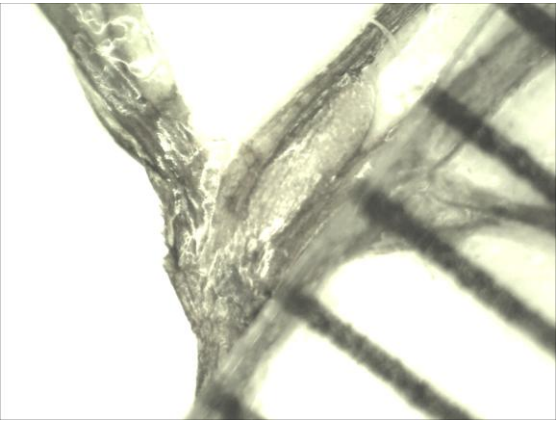




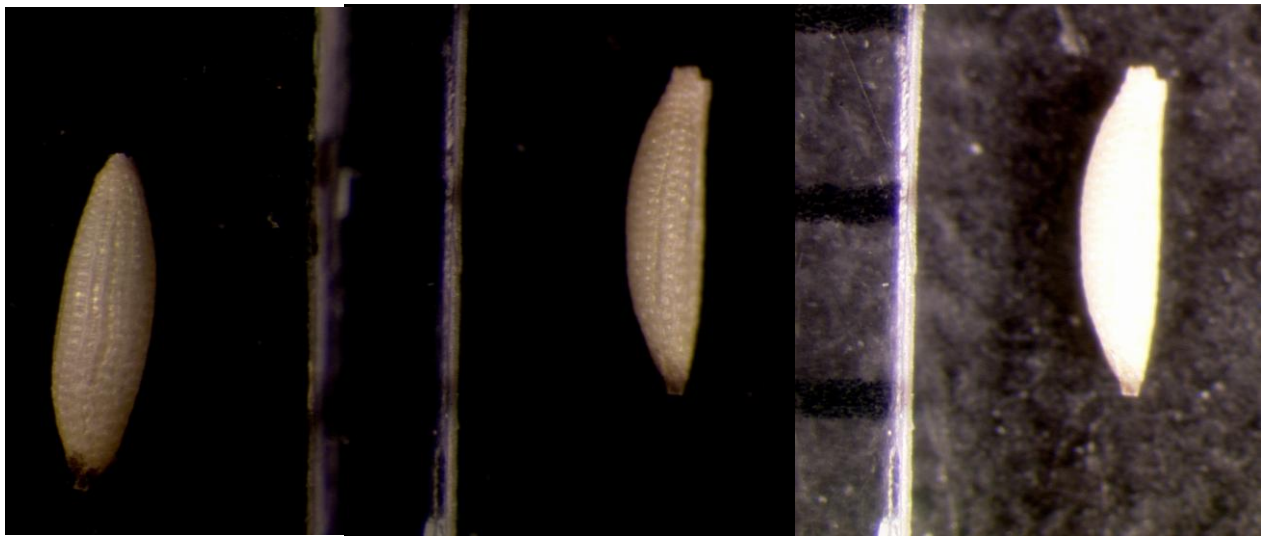
Najaguad198
South Dakota
No chloroplast
Universal: nrG9
R/S: fail no amp

cG3
2 out of 4 have seed

Najaguad124_glue
Texas
Universal: nr9
R/S: nrC nrF



Najaguad062
Texas
Universal: nrG5
R/S: nrC

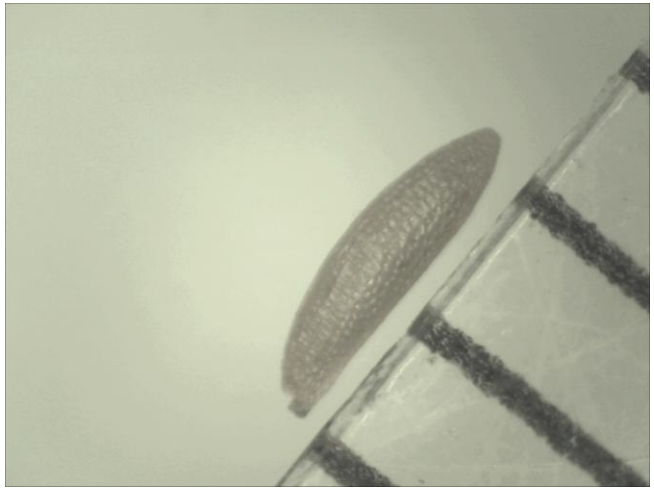


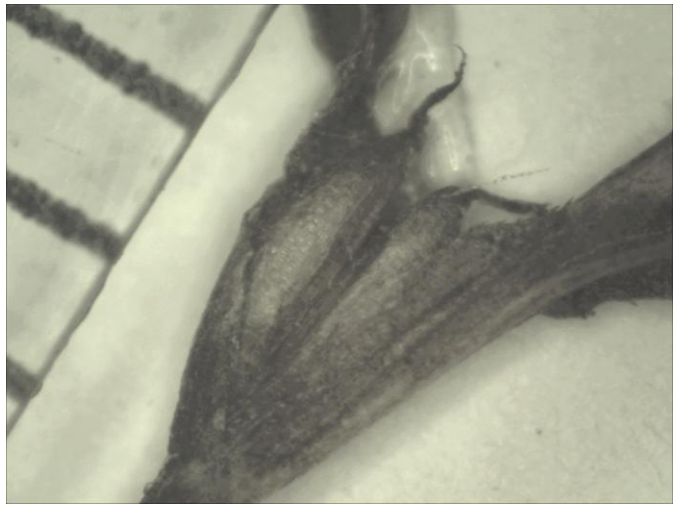
cG4
Only 1 specimen – no seed

Najaguad074
Kansas
Universal: nrG9 nrG10
R/S: amp guad not flex

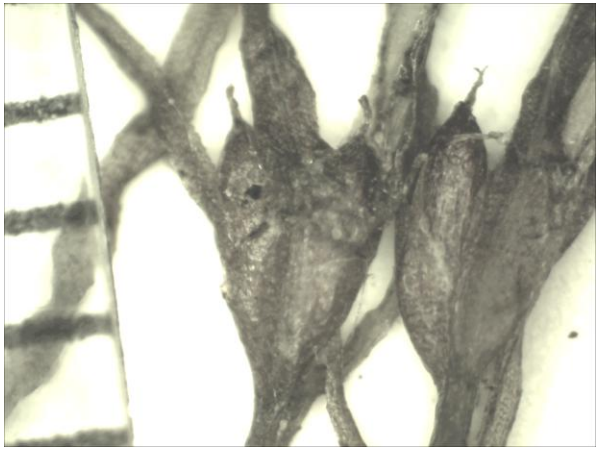
cG5

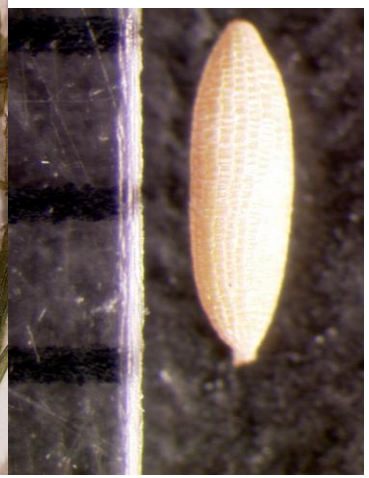
Najaguad133/134
Elephant Mountain_Texas
cG5
Universal: nrG2
R/S: nrC nrF
Many seed per node



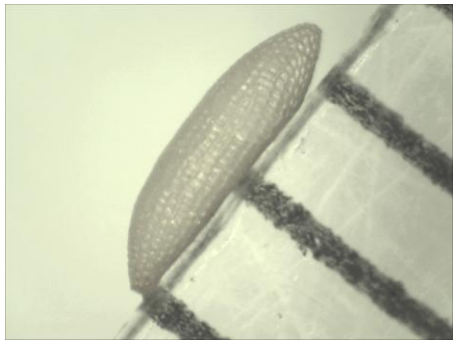


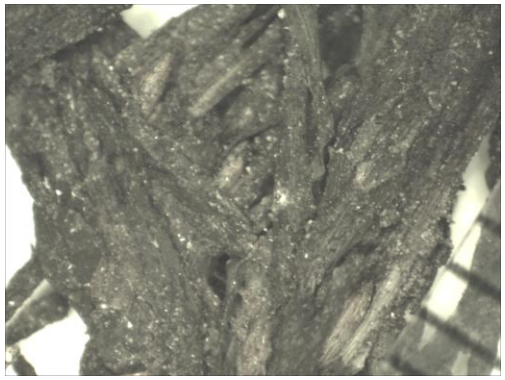
Najaguad132
Madera Creek_Texas
cG5
Universal: nrG2
R/S: nrC nrF
Many seed per node



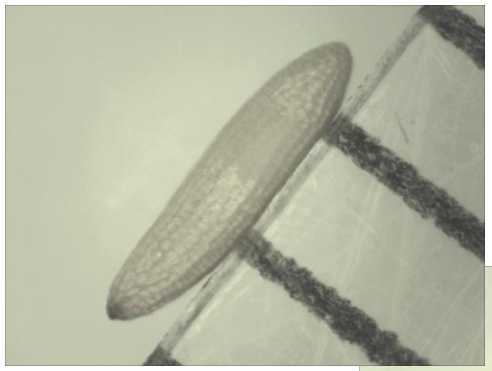


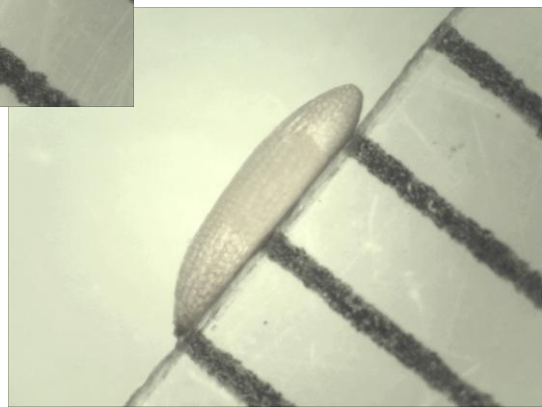
Hellquist_17190
Madera Creek _ Texas
Not extracted - many seed





Najaguad131
Madera Creek_Texas
Failed - many seed – 2 per node





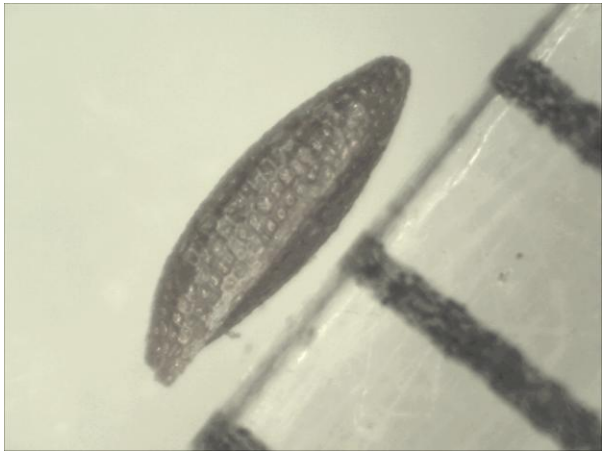
cG6
Only 1 specimen – no seed

Najaguad019
Florida
Universal: nrG1 nrG7
R/S: nrG1b

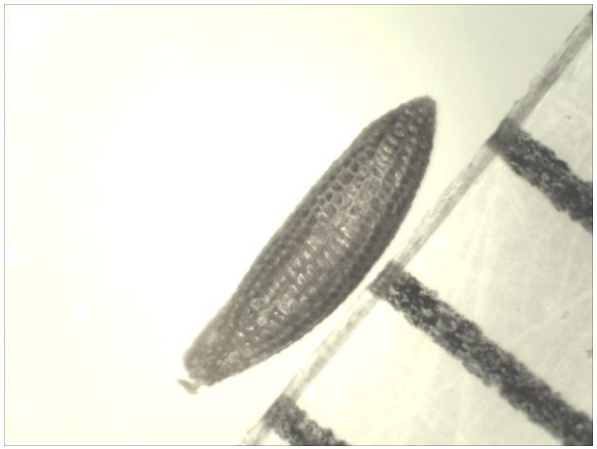
cG7

Najaflor006
Florida
cG7b
Universal: nrG1 nrG11
R/S: nrG1
Visible teeth
Only 2 seed

Najaflor004
Florida
No chloroplast region
Universal: nrG1
R/S: failed no amp
Visible teeth



Najaflor008
Florida
No chloroplast region
Universal: nrG1
R/S: failed no amp
Visible teeth



cG8
Only 1 specimen – no seed

Najaguad010
Florida
Universal: nrG1 nrG11
R/S: nrG1

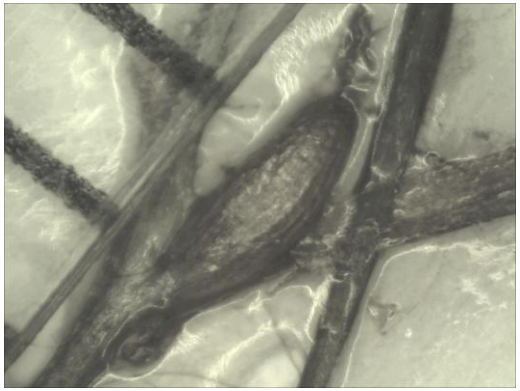
cG9
Only 1 specimen – no seed

Najaguad075
Missouri
Universal: nrG11
R/S: amp guad not flex

cG10
23:59 have seed

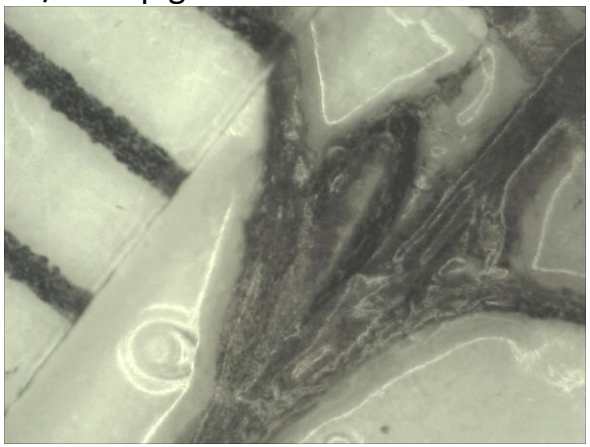
Najaguad084
Illinois
Universal: nrG8 nrG10
R/S: faint amp

Najaguad082_glue
Illinois
Universal: nrG8 nrG10
R/S: nrUn

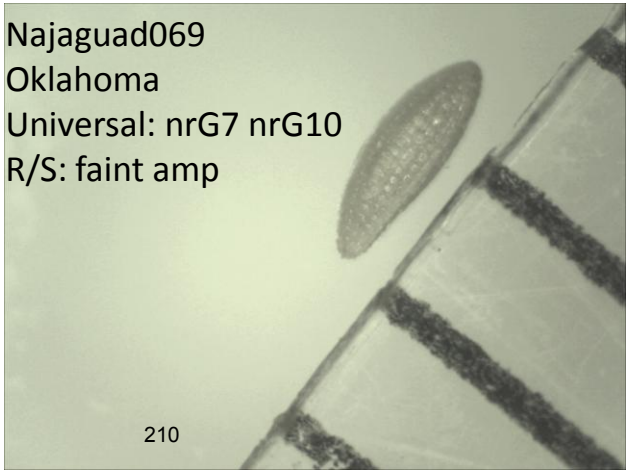


Najaguad085
Illinois
Universal: nrG8 nrG10
R/S: faint

Najaguad068_glue
Oklahoma
Universal: nrG7 nrG10
R/S: amp guad not flex

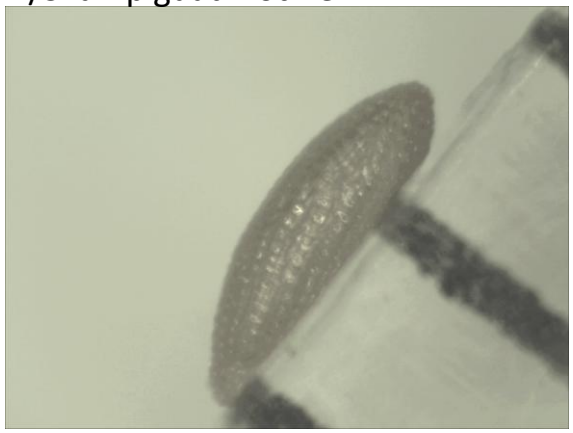


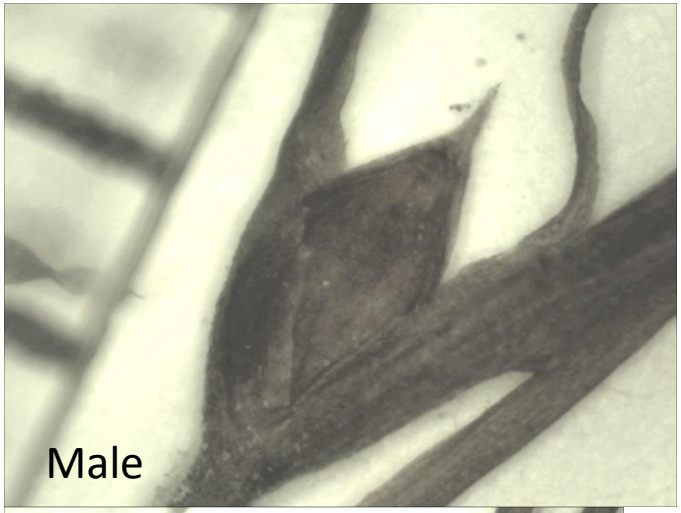
Najaguad069
Oklahoma
Universal: nrG7 nrG10
R/S: faint amp



cG10 cont.

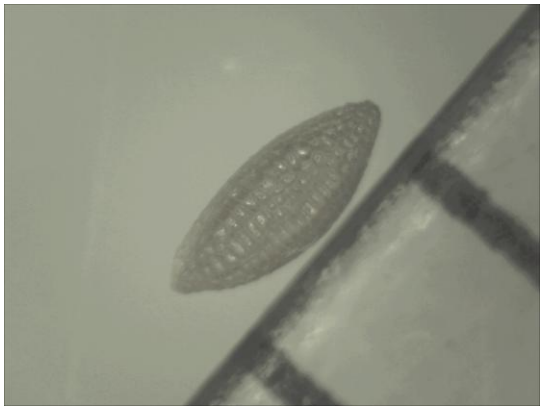
Najaguad078
Missouri
Universal: nrG8 nrG10
R/S: amp guad not flex



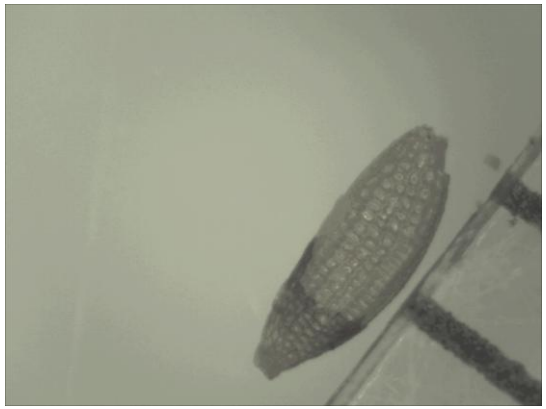


cG10 cont.

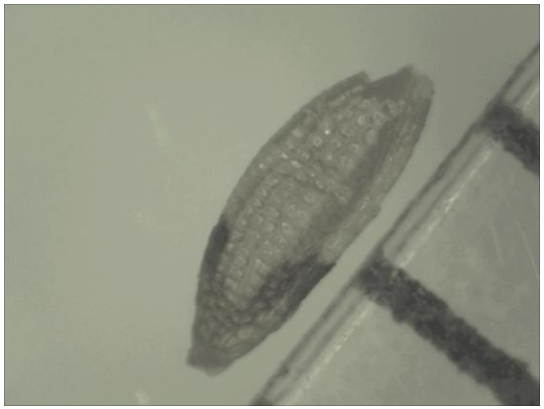
Najaguad171
Arizona
Universal: nrG7 nrG11
R/S: nrF nrC
Many seed



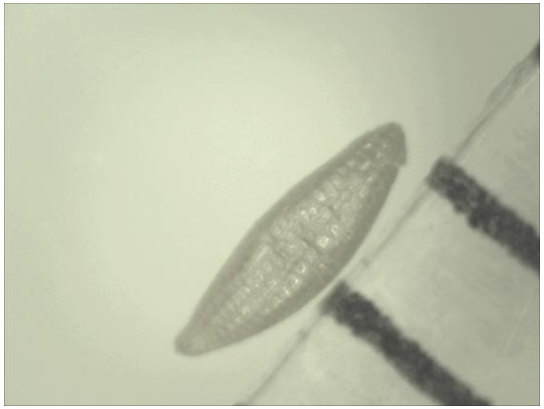
Najaguad172 Arizona
Universal: nrG11
R/S: nrC



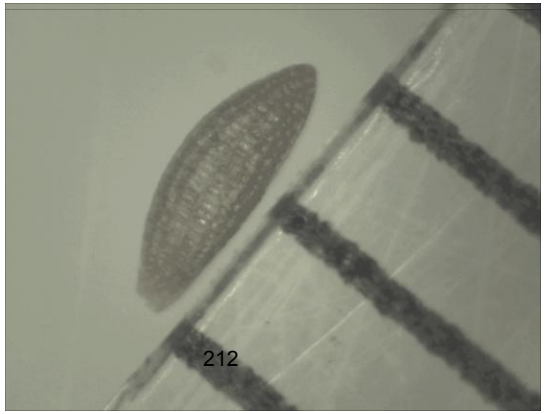
Najaguad172 has embryo



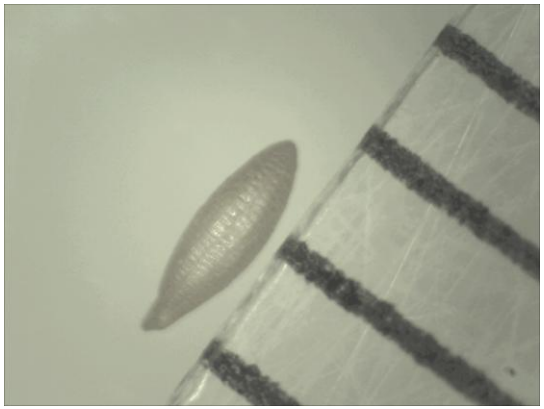
Najaguad086
Illinois
Universal: nrG9 nrG10
R/S: gel band
2 hollow_1 glue



Najaguad168
California
Universal: nrG4 nrG10
R/S: Faint amp
Many seed but many look inviable

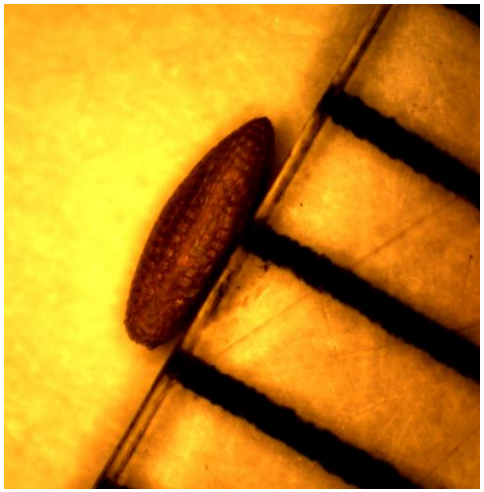


Najaguad096
Nebraska
Universal: nrG9 nrG10
R/S: amp guad not flex
Many seed

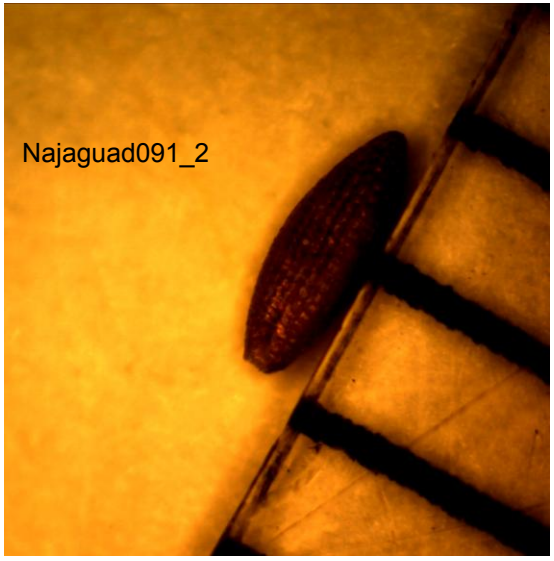


cG10 cont.

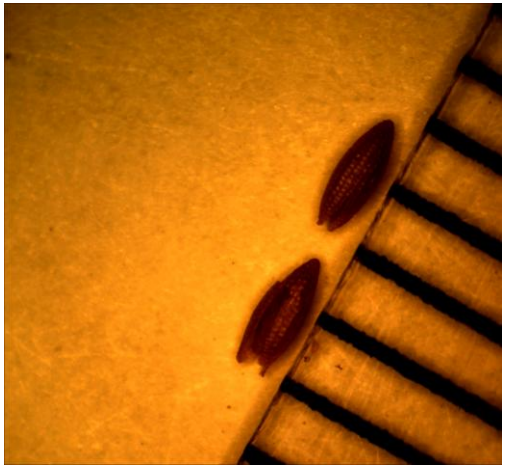
Najaguad091
Iowa
Universal: nrG9 nrG10
R/S: amp guad not flex



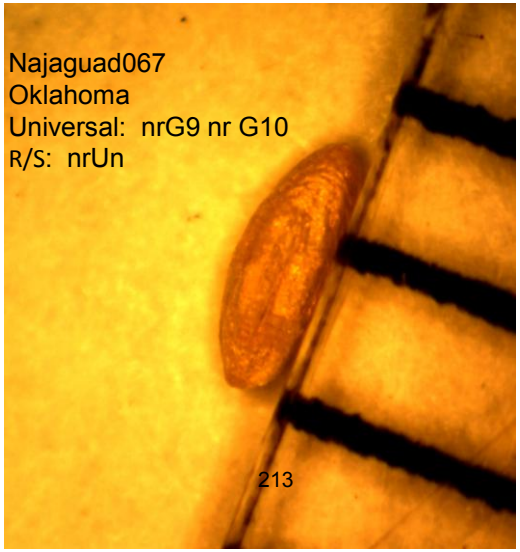
Najaguad091_2



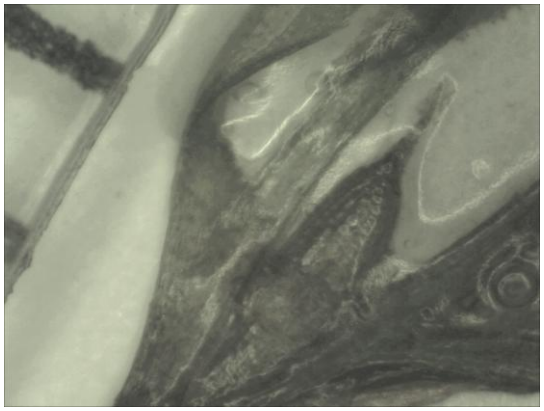
Najaguad032
Iowa
Universal: nrG10
R/S: nrC nrF



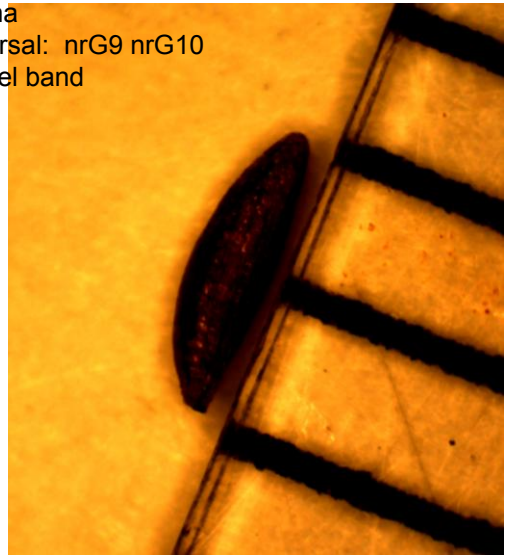
Najaguad067
Oklahoma
Universal: nrG9 nr G10
R/S: nrUn



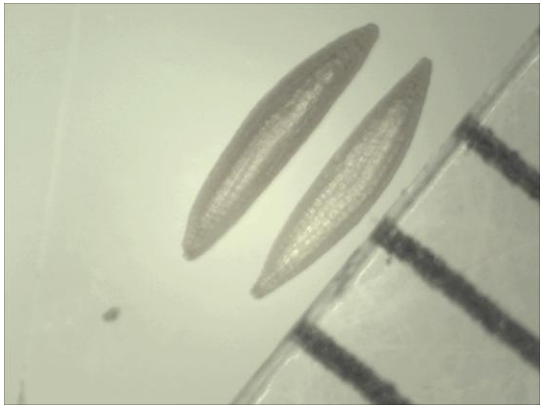
Najaguad079_glue
Missouri
Universal: nrG8 nrG10
R/S: amp guad not flex

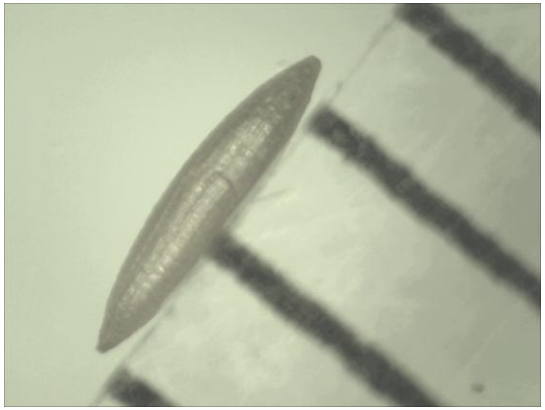


Najaguad054
Indiana
Universal: nrG9 nrG10
R/S: gel band

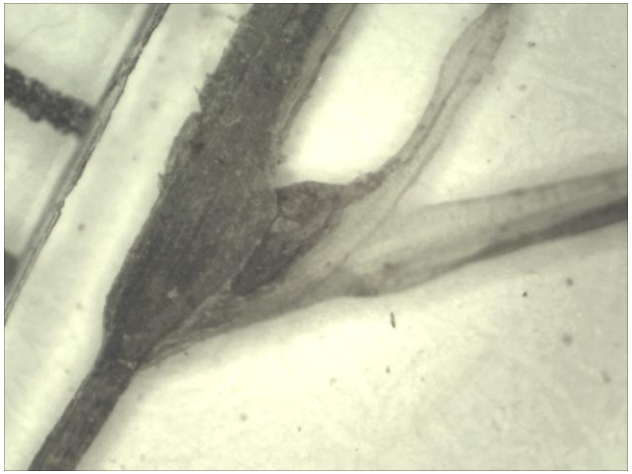


cG10 cont.





Najaguad223
Kentucky
Universal: nrG9
R/S: nrU2
Many shiny seed

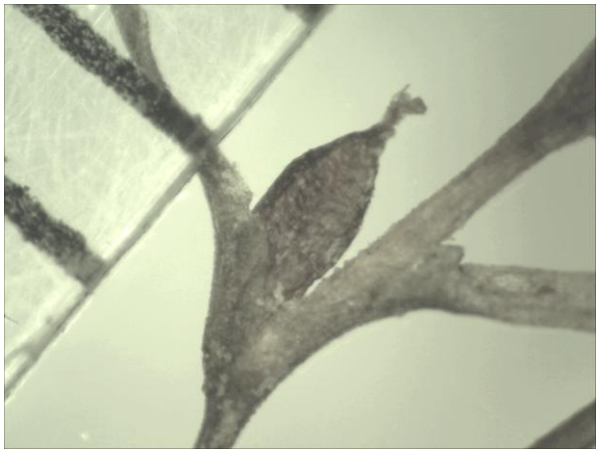


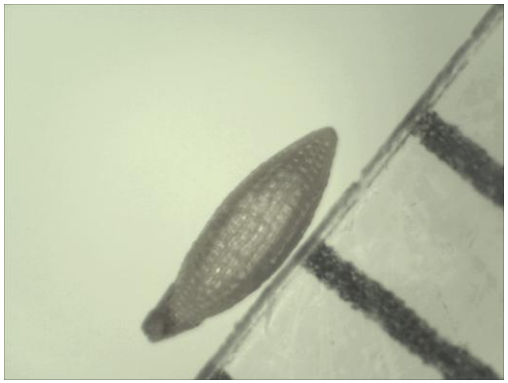
Male flowers

Male flowers

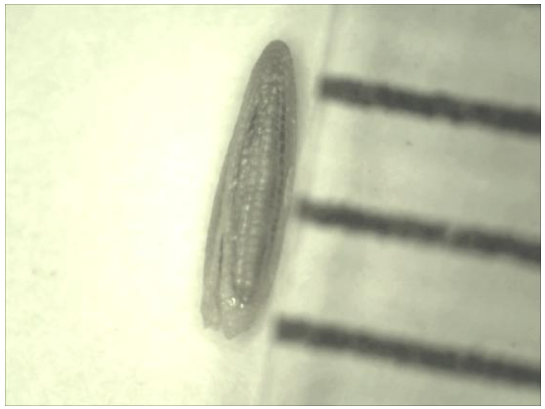
cG10 cont.

Najaguad148 and 245 Klaus 8b1/8b2
Illinois
Universal: nrG8
Universal: nrG8 nrG10
R/S: amp guad not flex
Mixed herbarium sheet
Other guad morphology on herbarium sheeet with long narrow seeds like 2 samples over

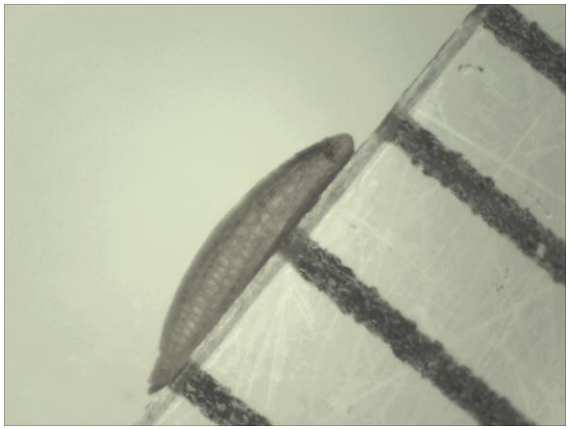




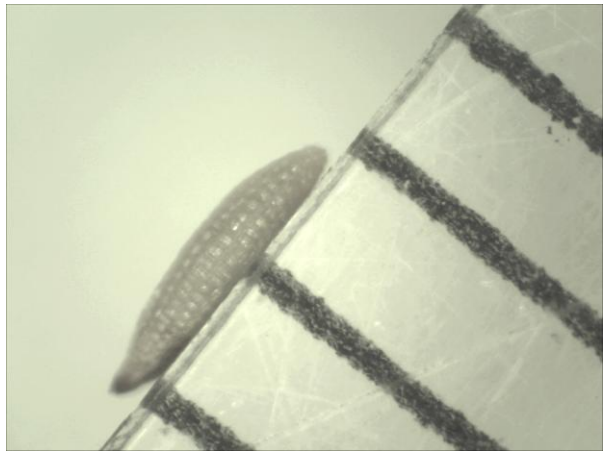
Najaguad207
California
No chloroplast region
Universal: nrG10 nrG11
R/S: amp guad not flex
nrG10 nrG11



Najaguad236
West Virginia
Universal: nrG3 nrG11
R/S: amp guad not flex

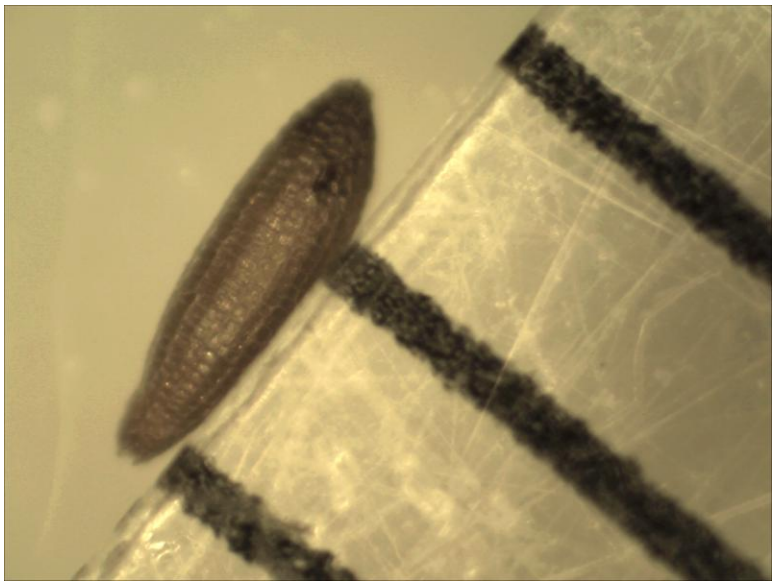


Najaguad237
Kansas
Universal: nrG7 nrG10
R/S: amp guad not flex
Lots seed and males



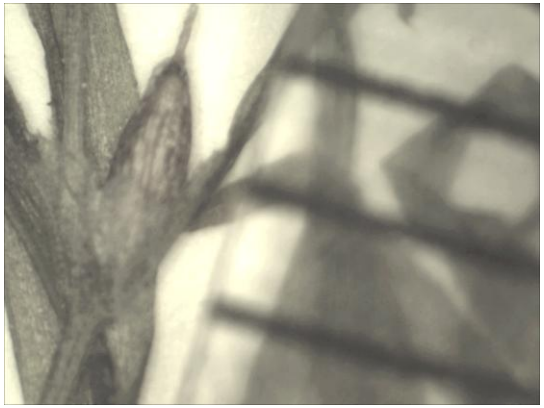
cG10 cont.

Najaguad043
OK
Universal: nrG7 nrG10
R/S: nrUn

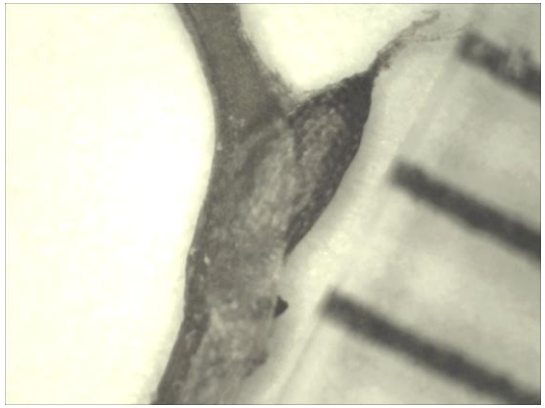


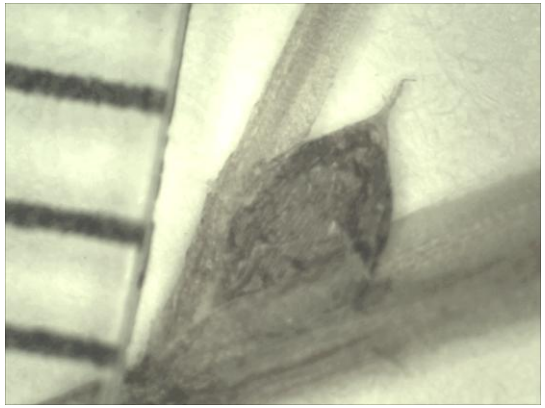
Najaguad150
AZ
Universal: nrG7 nrG11
R/S: nrF nrC
(most seed hollow)

Najaguad209
California
No chloroplast
Universal: nrG9
R/S: amp guad not flex
Many viable seed



Najaguad213
California
No chloroplast
Universal: nrG9 nrG111
R/S: Faint





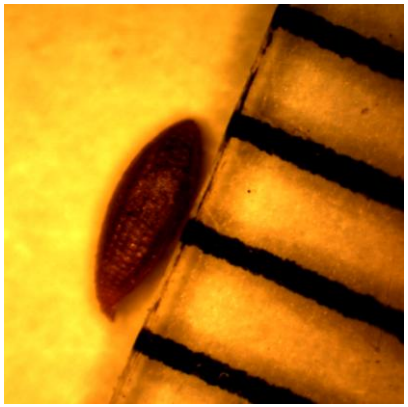
cG11
only 2 specimens – no seed

Najaguad063
Texas
Universal: nrG8
R/S: nrUn

Najaguad072
Kansas
Universal: nrG9 nrG10
R/S: amp guad not flex

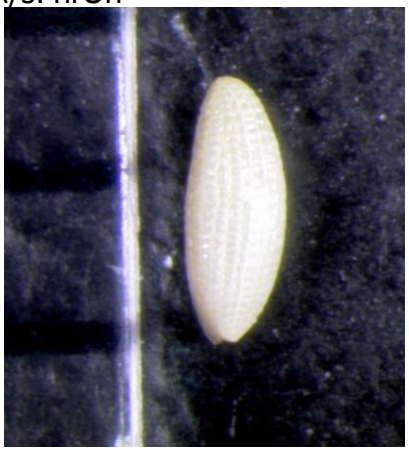
cG12 (6:8 have seed)

Najaguad178
California
Universal: nrG3
R/S: Gel band

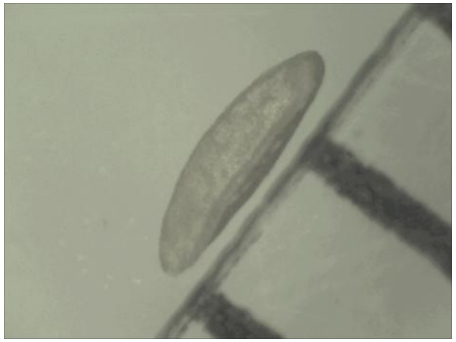


Najaguad071
Kansas
Universal: nrG8
R/S: amp guad not flex

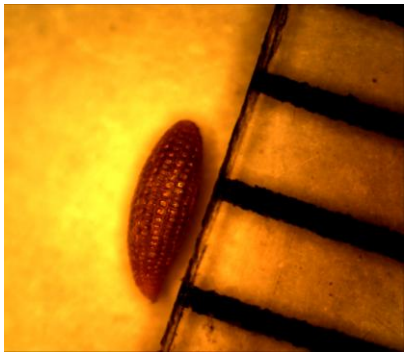
Najaguad175
Arizona
Universal: nrG9 nrG11
R/S: nrUn



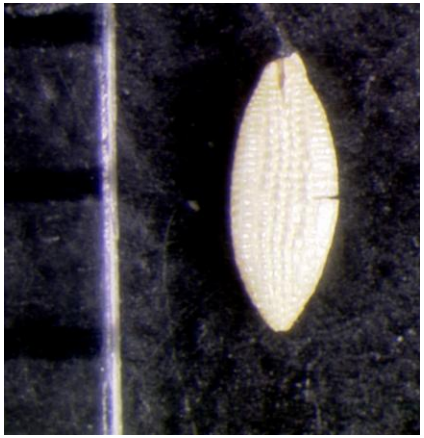
Najaguad176
California
Universal: nrG9 nrG11
R/S: nrUn
2 seed hollow



Najaguad031
Iowa
Universal: nrG9 nrG10
R/S: nrC

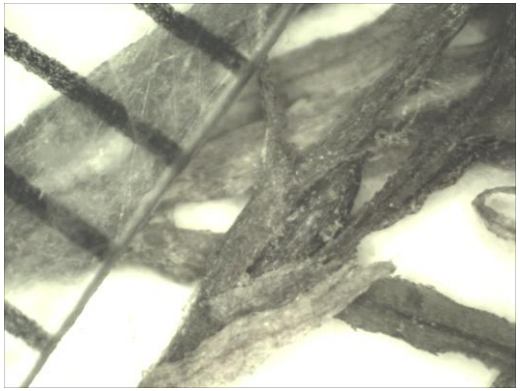


Najaguad170
New Mexico
Universal: nrG7 nrG11
R/S: nrC



cG13
2 out of 5 have seed

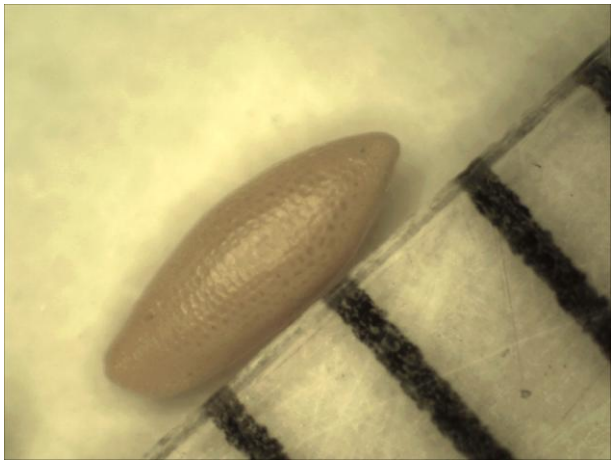
Najaguad214
California
Universal: nrG3
R/S: nrUn
Lots seed-CDA herbarium



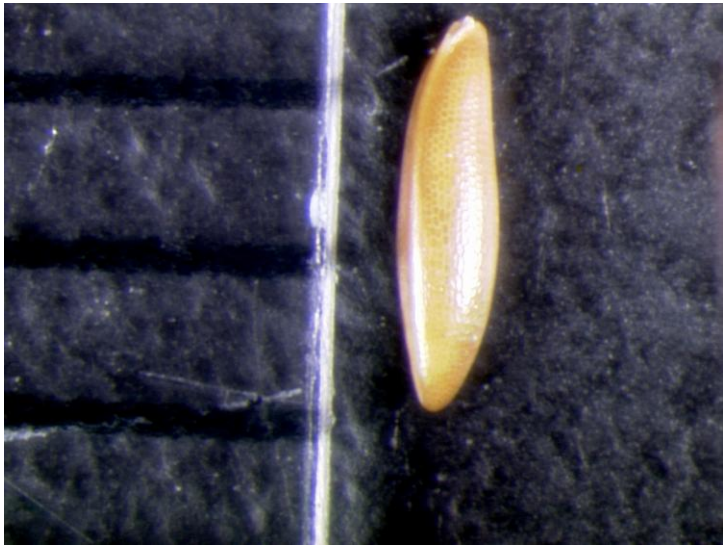
Najaguad169
California
Universal: nrG4 nrG10
R/S: fail no amp

cG14

Najaoliv238 (isolectotype)
Minnesota
Universal: nrG14
R/S: nrF

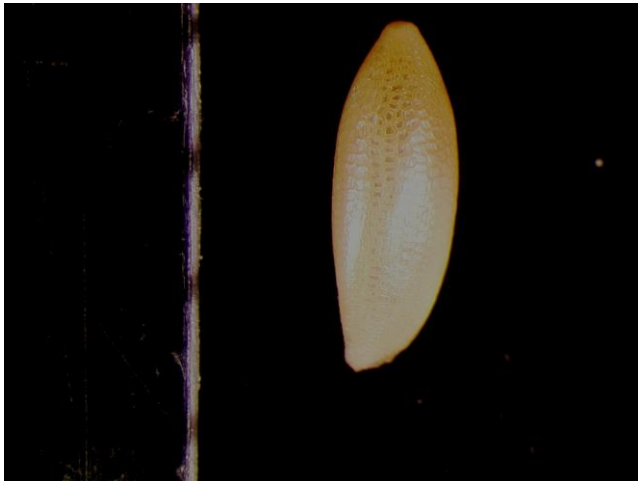


Najaoliv241
Minnesota
Universal: nrG14
R/S: fail no amp



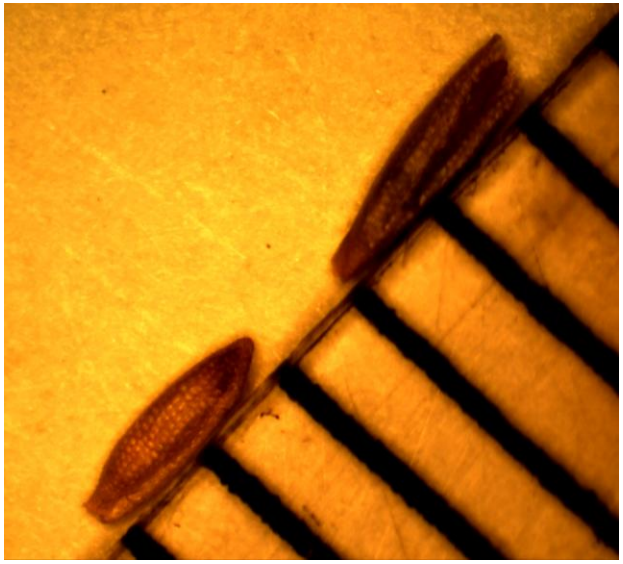
Najaoliv232
(isolectotype) Minnesota
No chloroplast region
Universal: nrG14
R/S: fail no amp

Guad238 - isolectotype
Guad232 - isolectotype

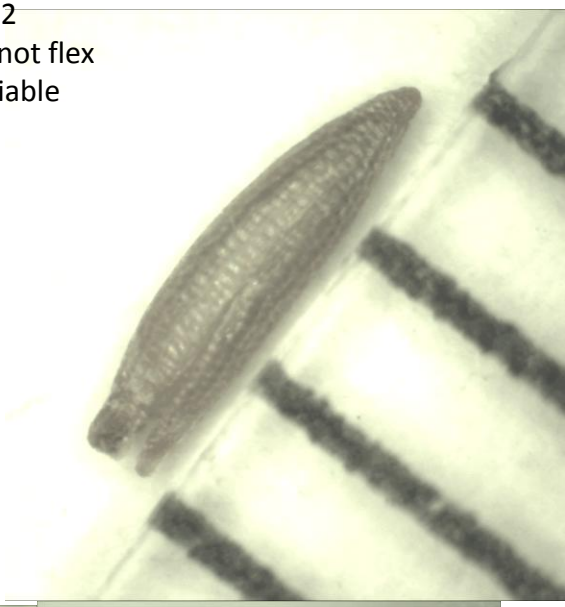


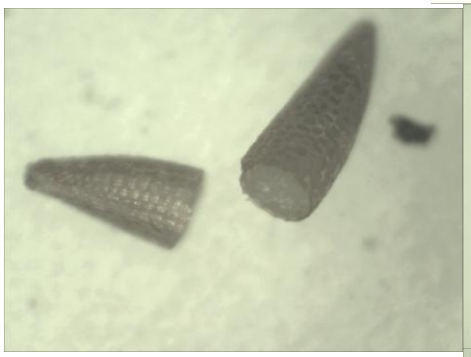
cG15 – 2 out of 15 have seed

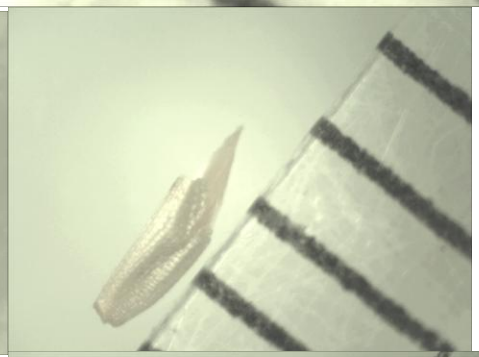
Najaguad006
Bantam Lk, Connecticut
Universal: nrG12
R/S: failed no amp

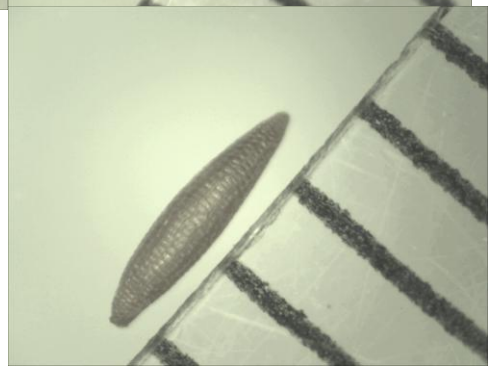


Najaguad251
Bantam Lk Connecticut
Universal: nrG12
R/S: amp guad not flex
Embryo looks viable



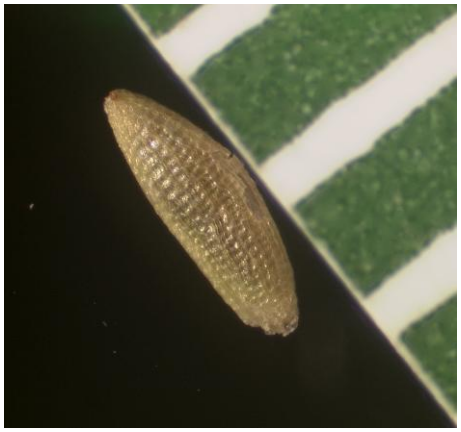




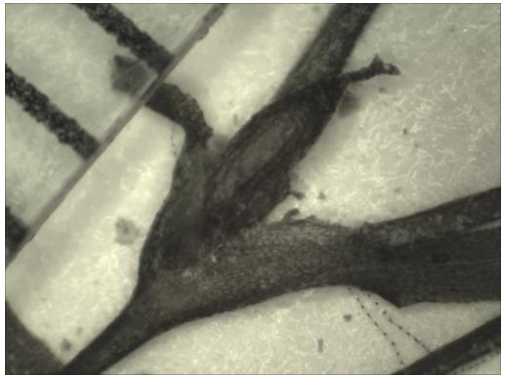


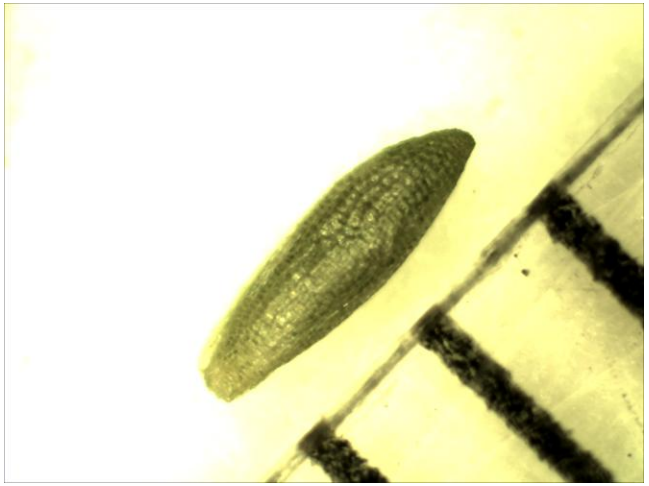
cG16
These are the only 2 accessions from this haplotype with seed

Najaflor001
South Carolina
cG16
Universal: nrG1 nrG13
R/S: nrG1
Anther: 1 locule



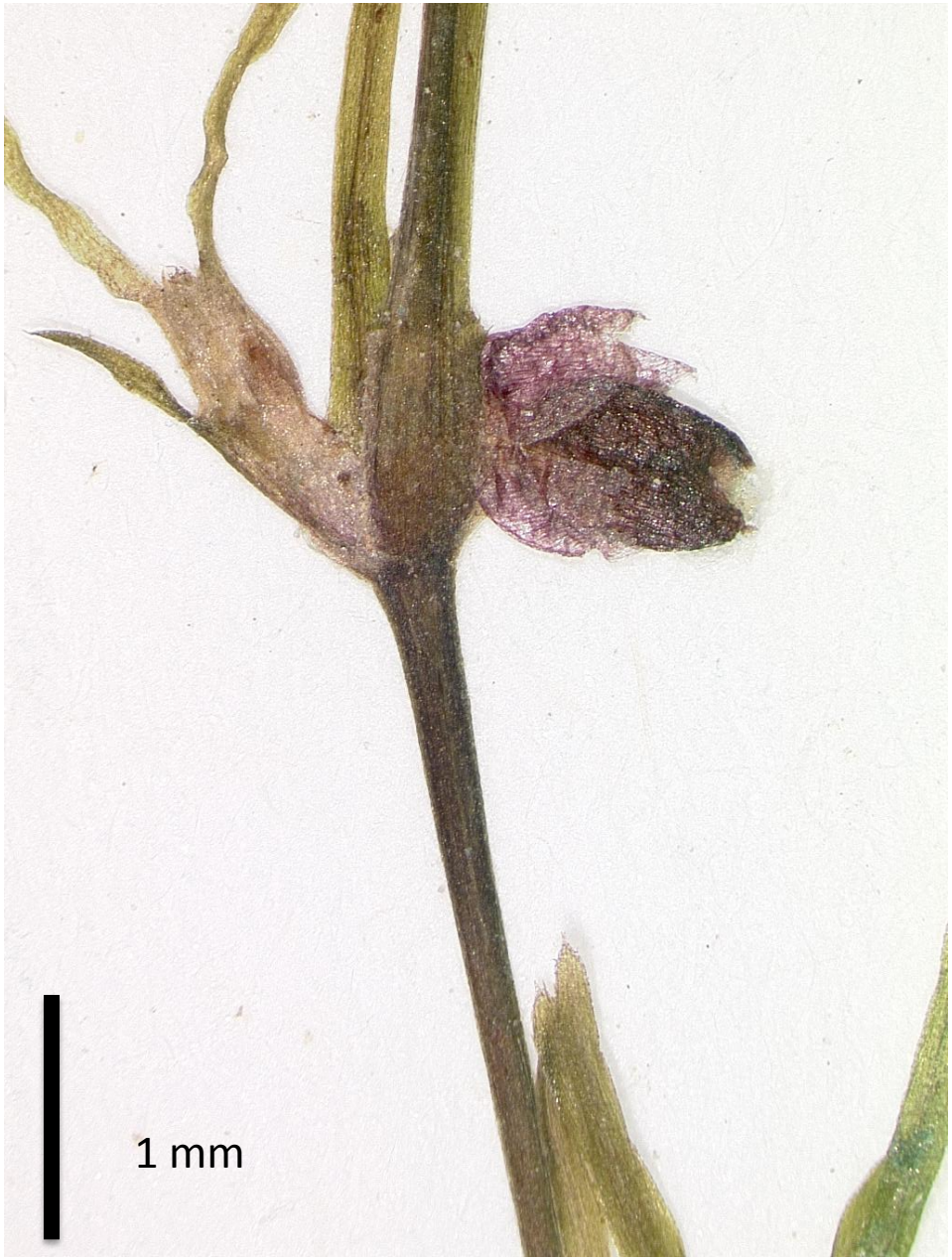
Najaguad114
Ohio
cG16
Universal: nrG12
R/S: nrC
3 seed – viable looking with embryos



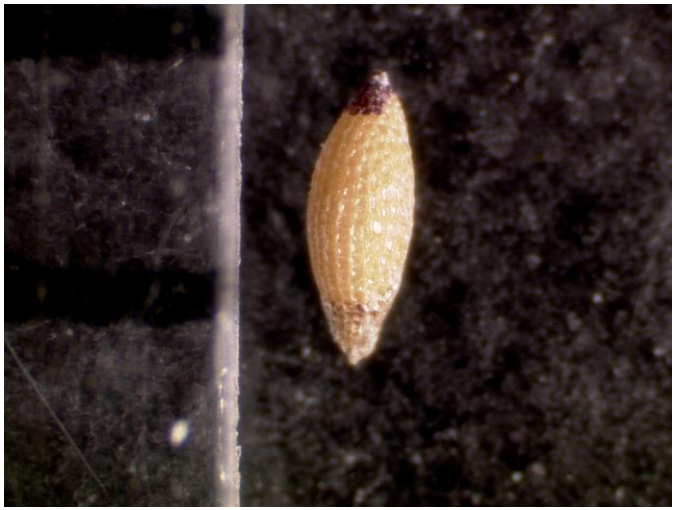


Najaguad078
(male flowers)

Najaguad078
Male flower
Dehiscing



Najawrig002



Appendix H

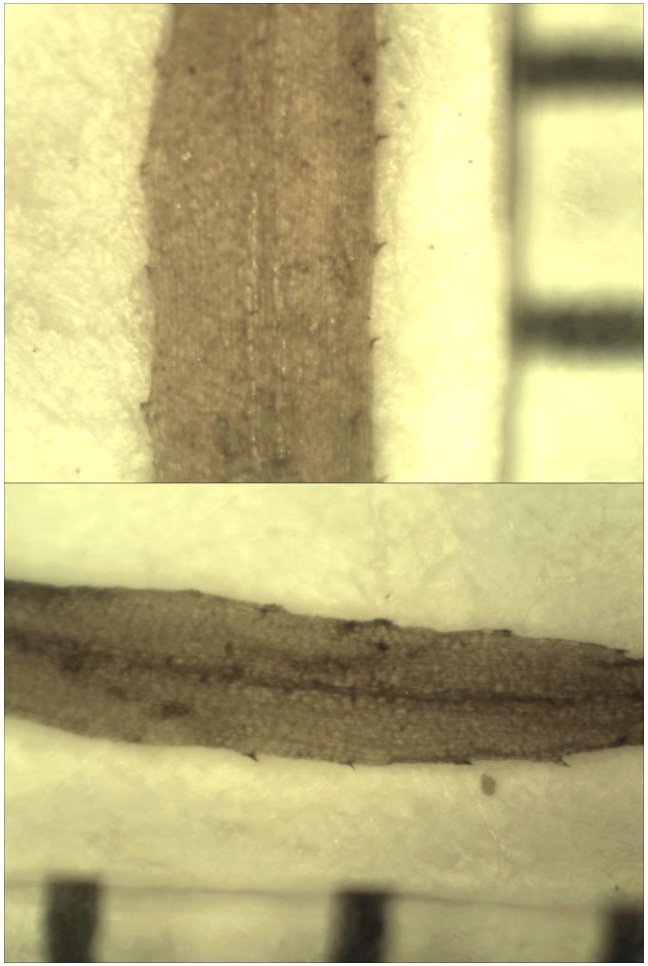
Leaf images

Herbarium specimens

cG1

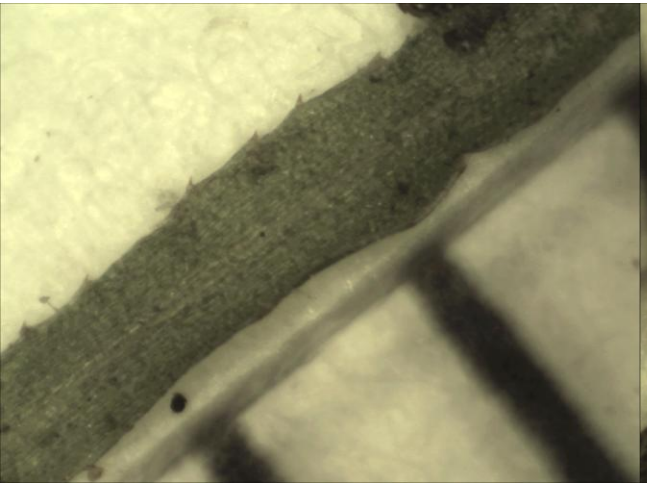
Najaguad255
Honduras

Najaguad146
(no chloroplast loci)
Costa Rica

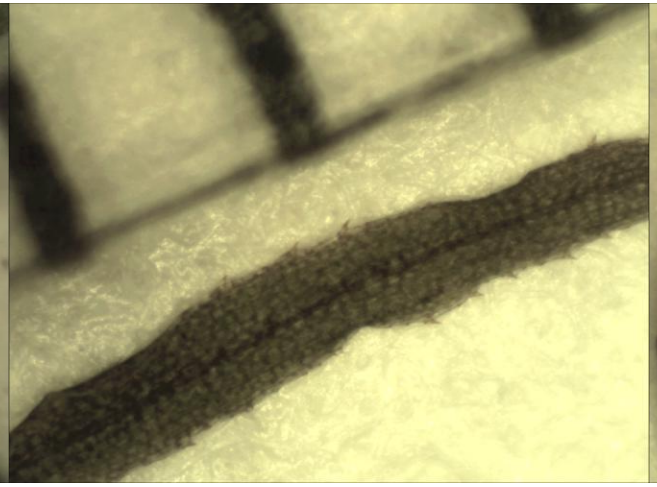


cG2

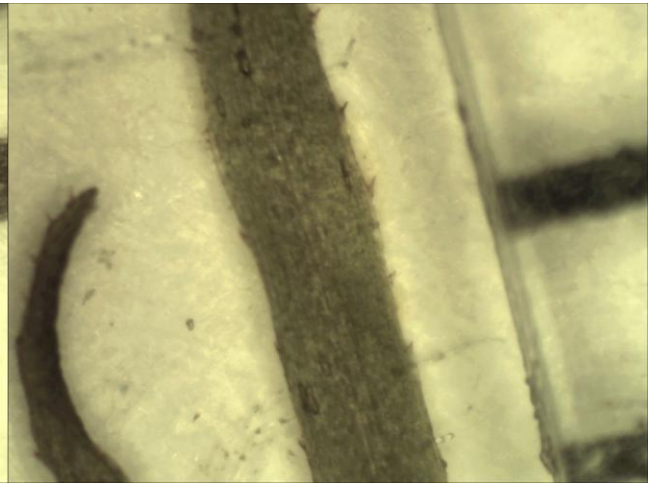
Najaguad244



Najaguad095

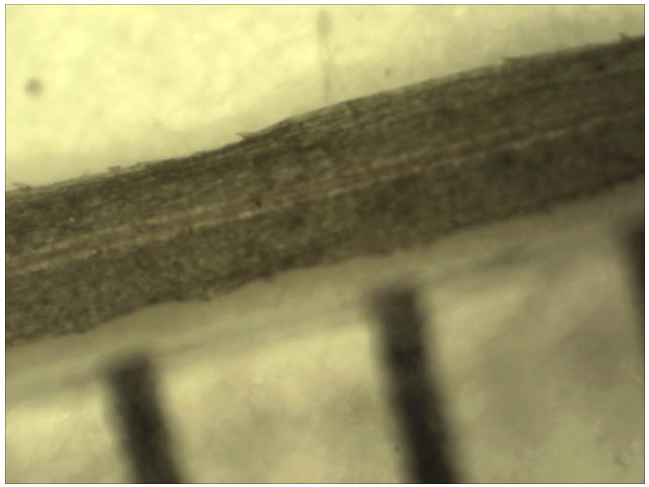


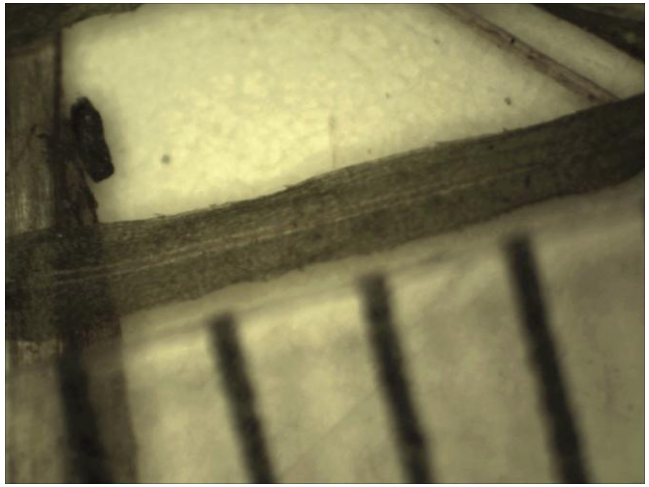
Najaguad094



cG2

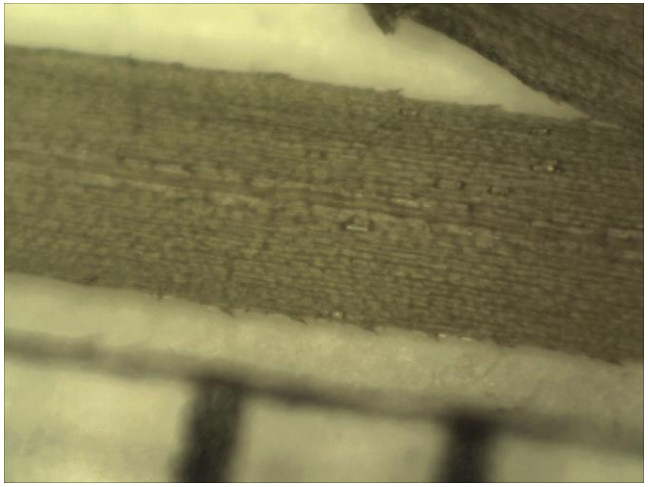
Najaguad196/197

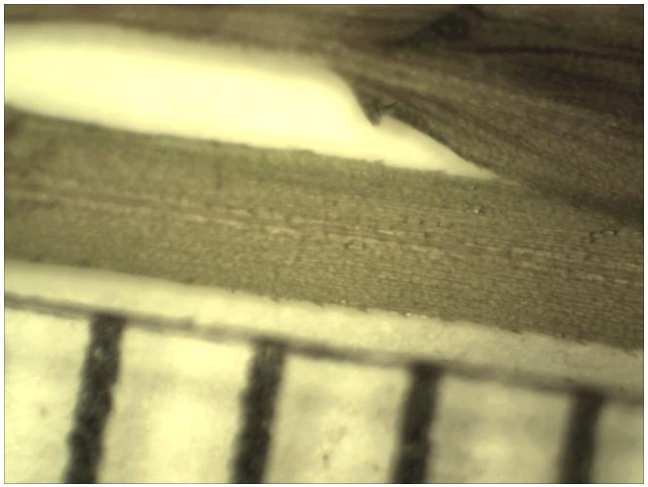




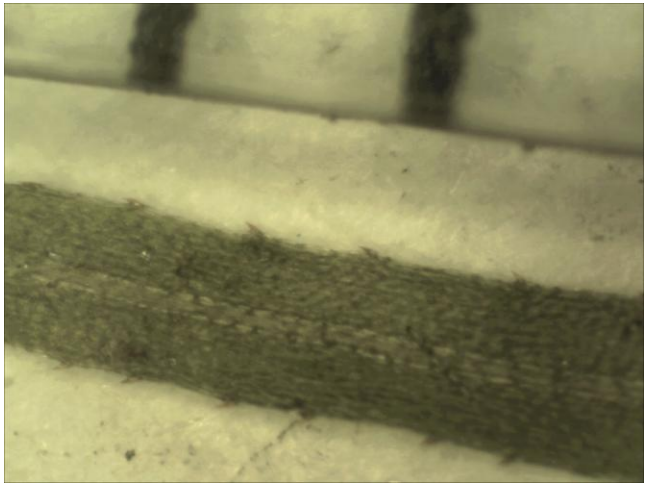
cG3

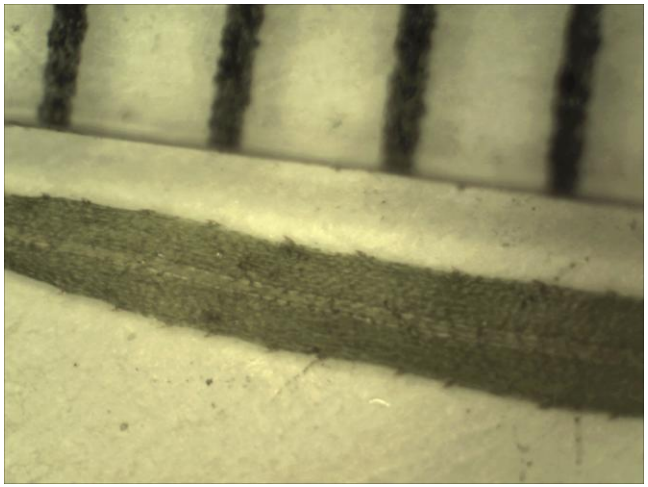
Najaguad062





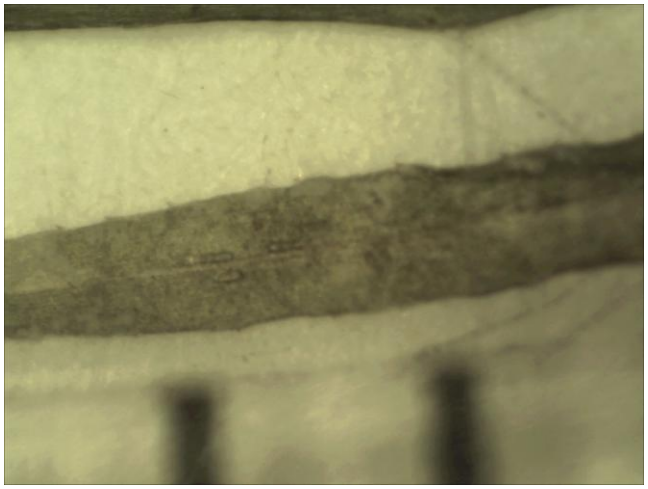
Najaguad065

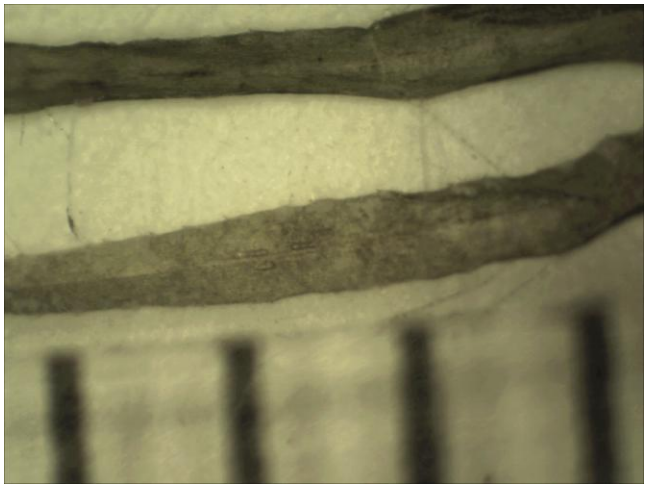




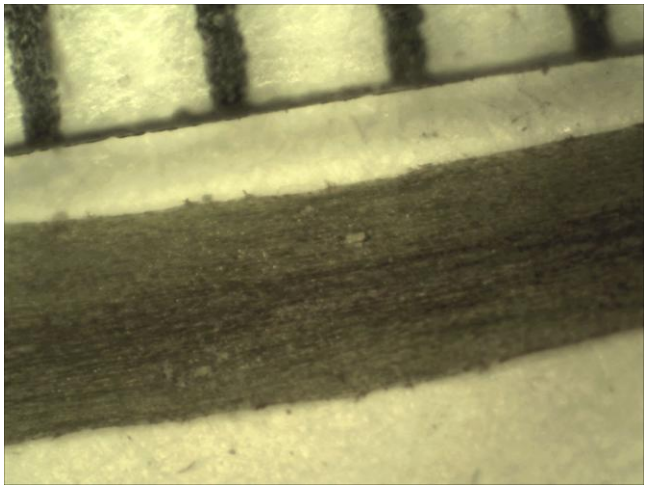
cG3

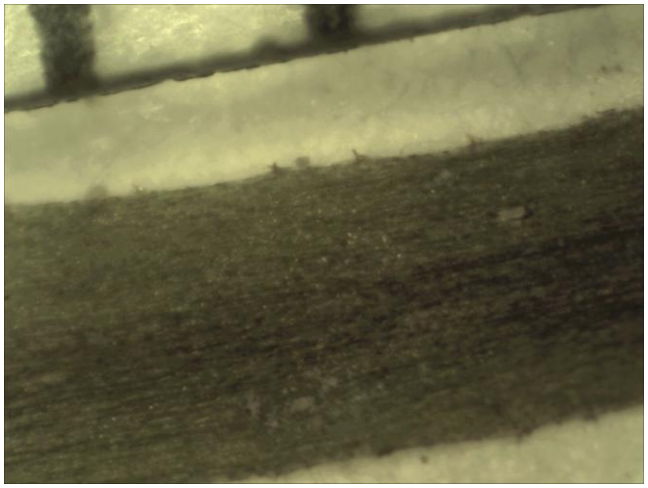
Najaguad124





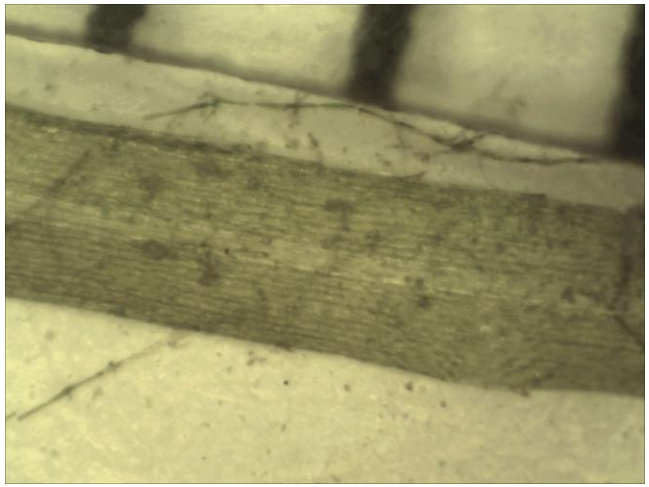
Najaguad177

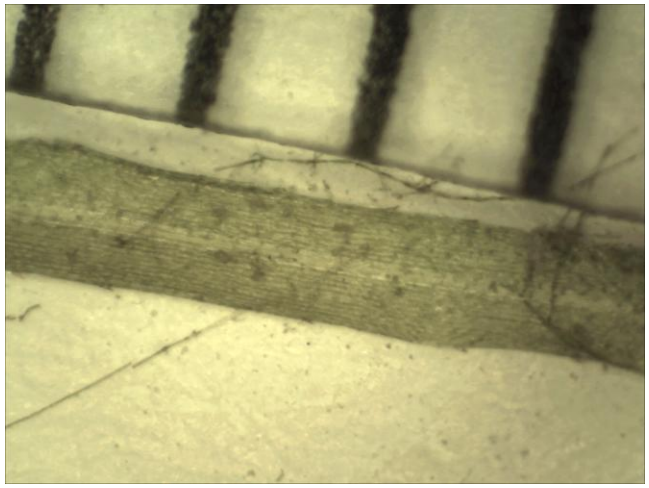




cG4

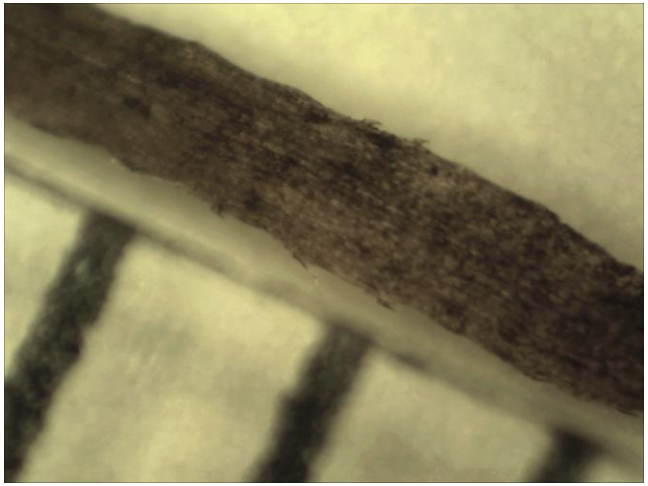
Najaguad074

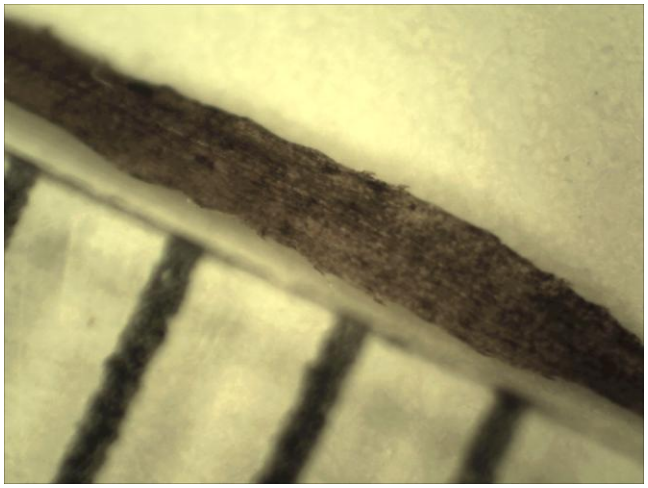




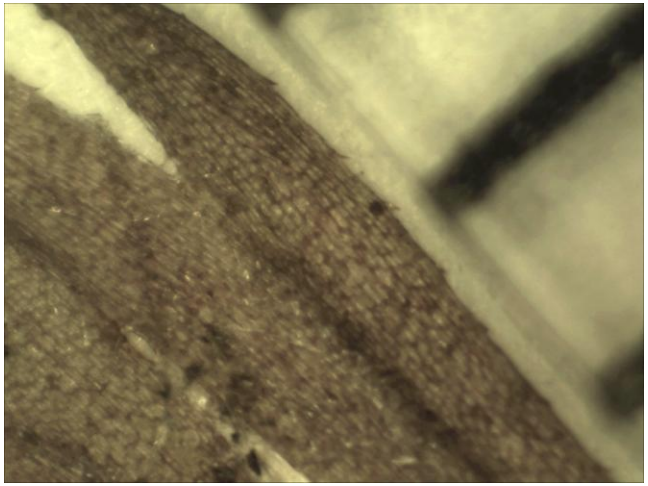
cG5

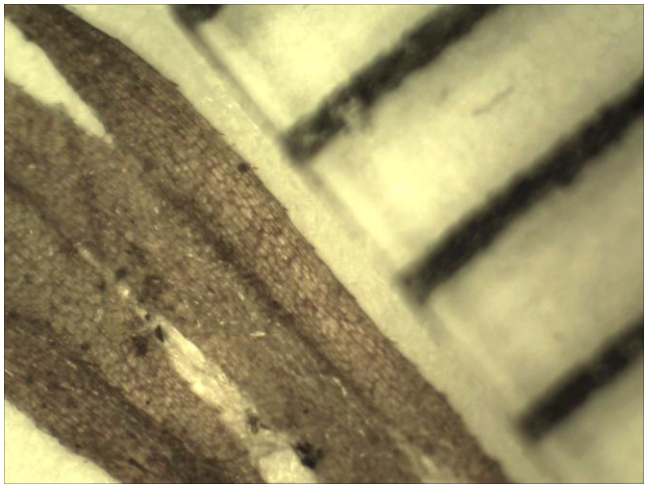
Najaguad133_134





Najaguad132

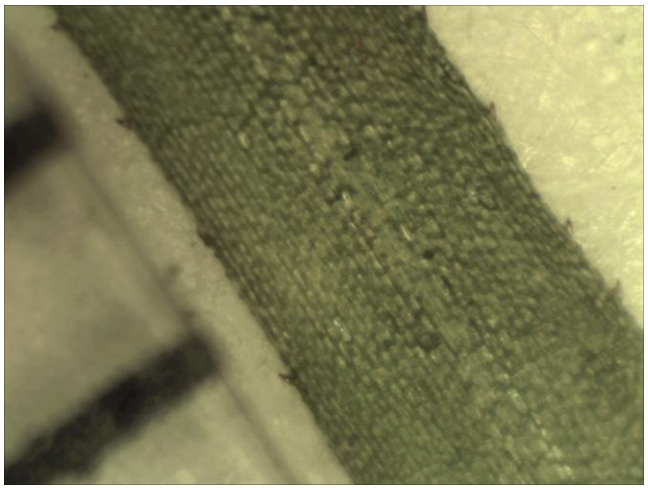


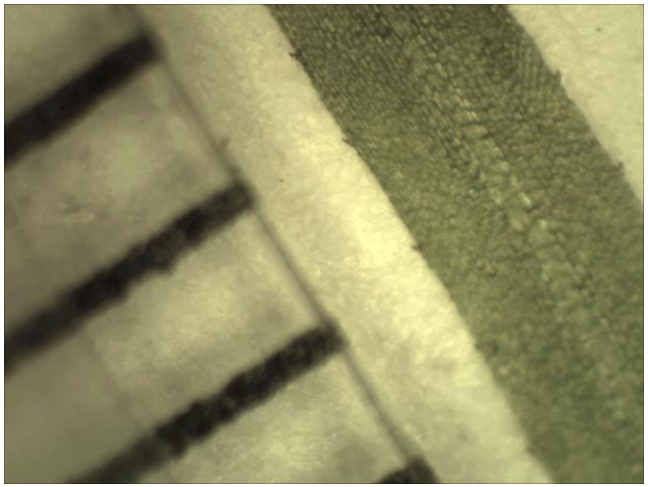


cG6

'floridana' clade

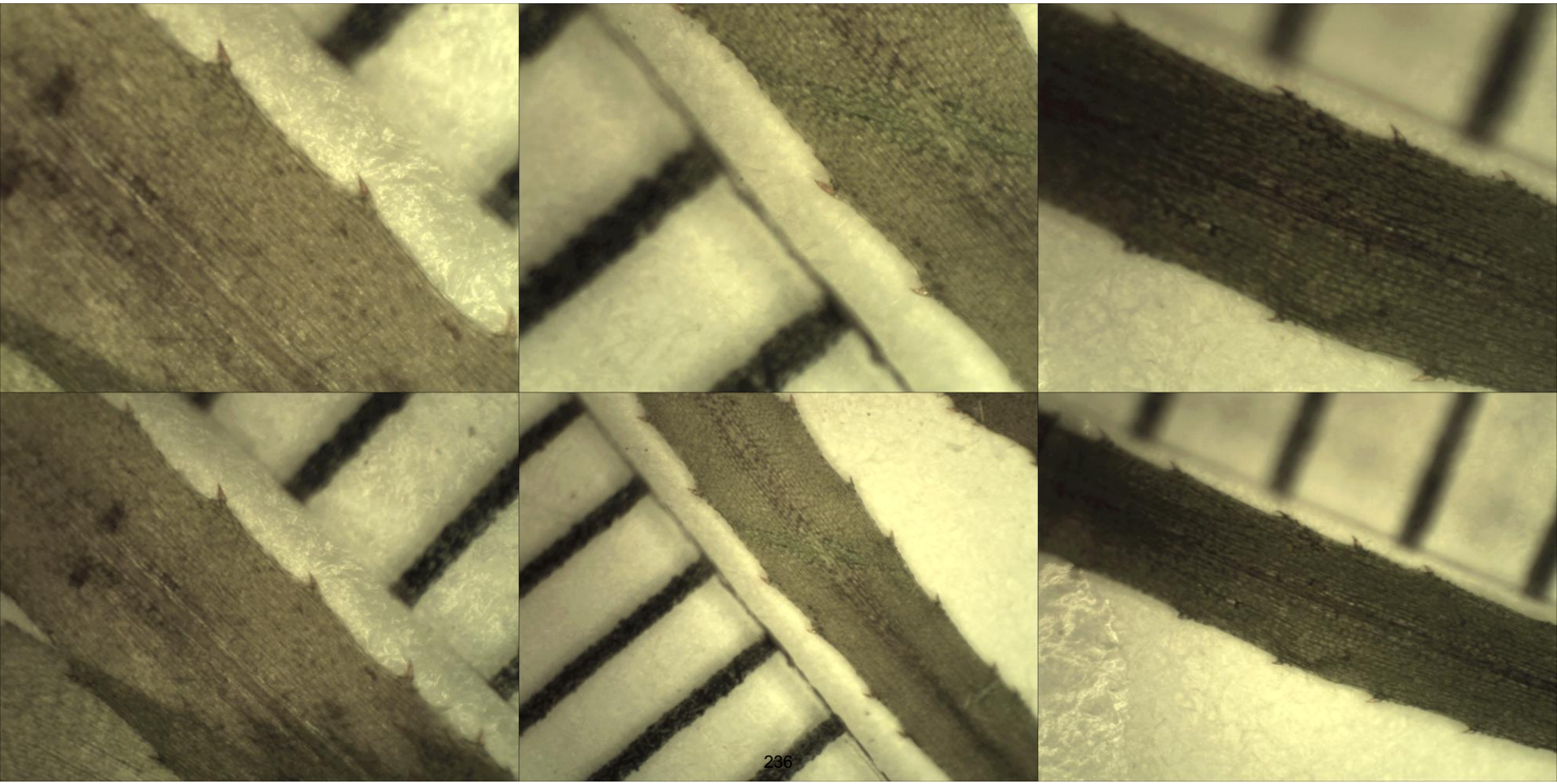
Najaguad019
Note: marginal teeth not exserted





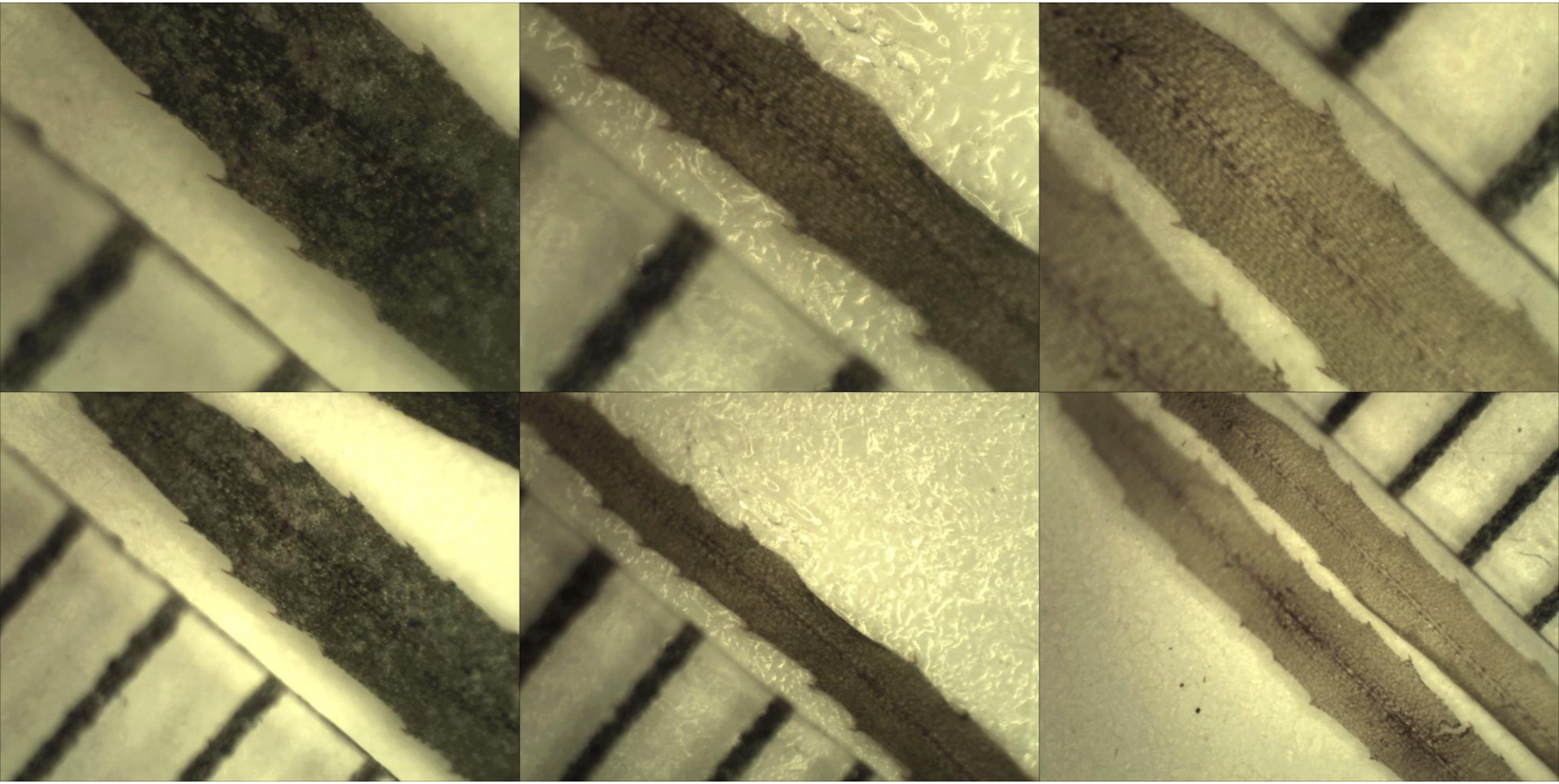
cG7
'floridana' clade

Najaflor003 Najaflor002 Najaflor007



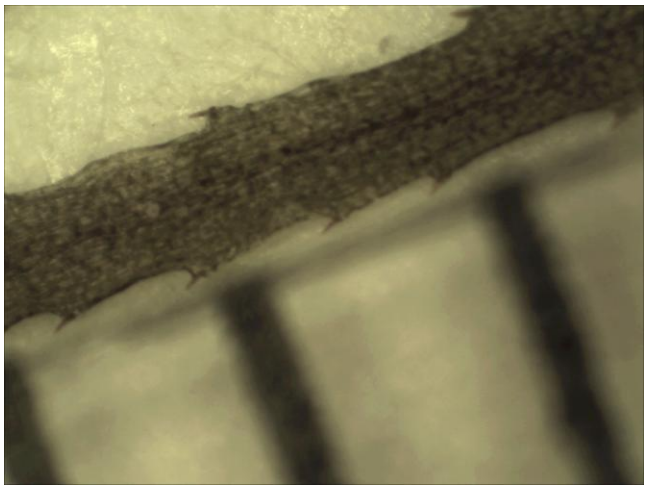
cG7
‘floridana’ clade

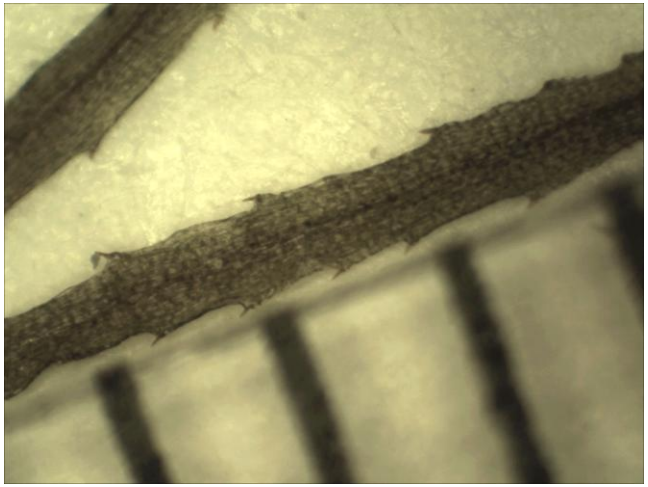
Najaflor016 Najaflor015 Najaflor006



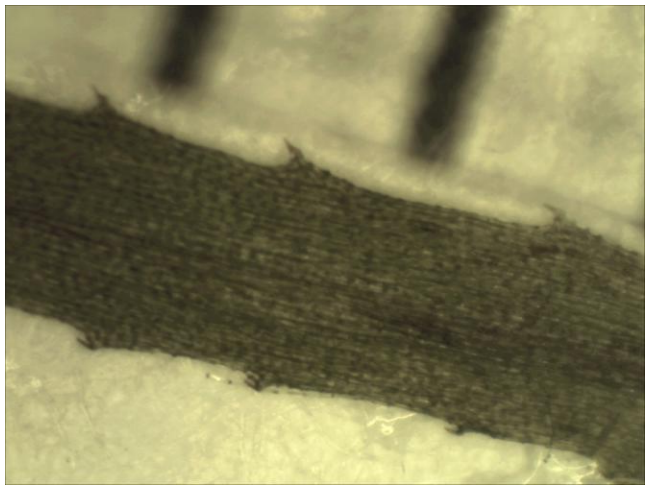
cG7
'floridana' clade

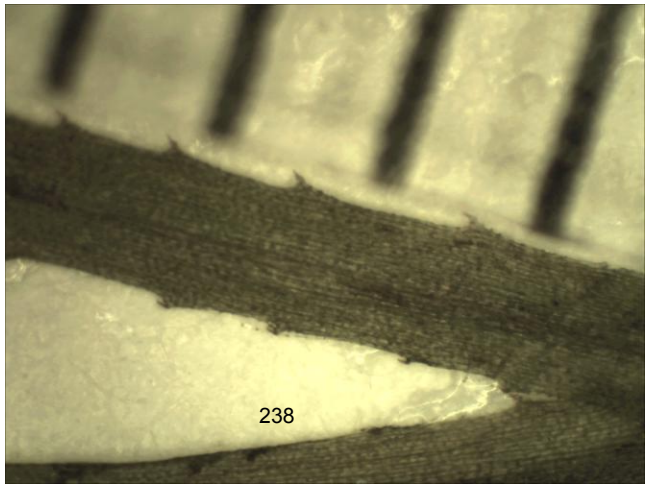
Najaflor009



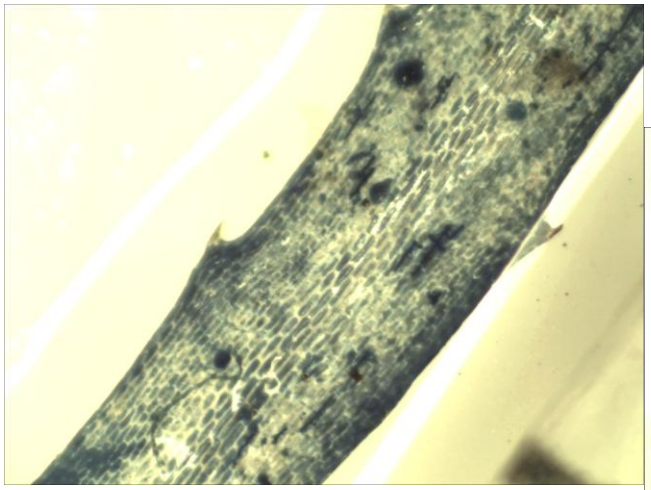


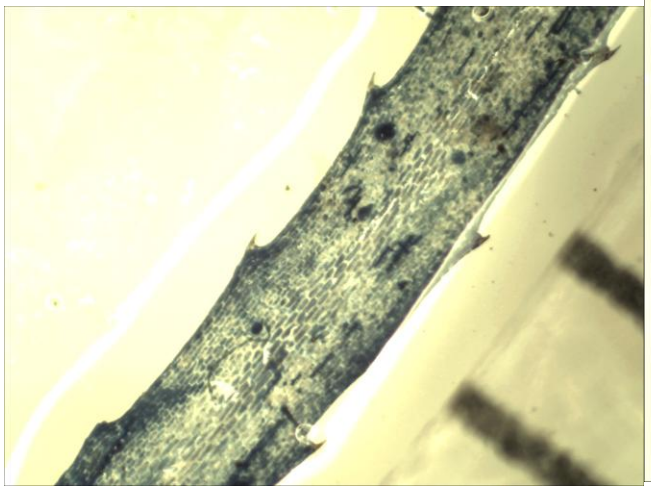
Najaflor014

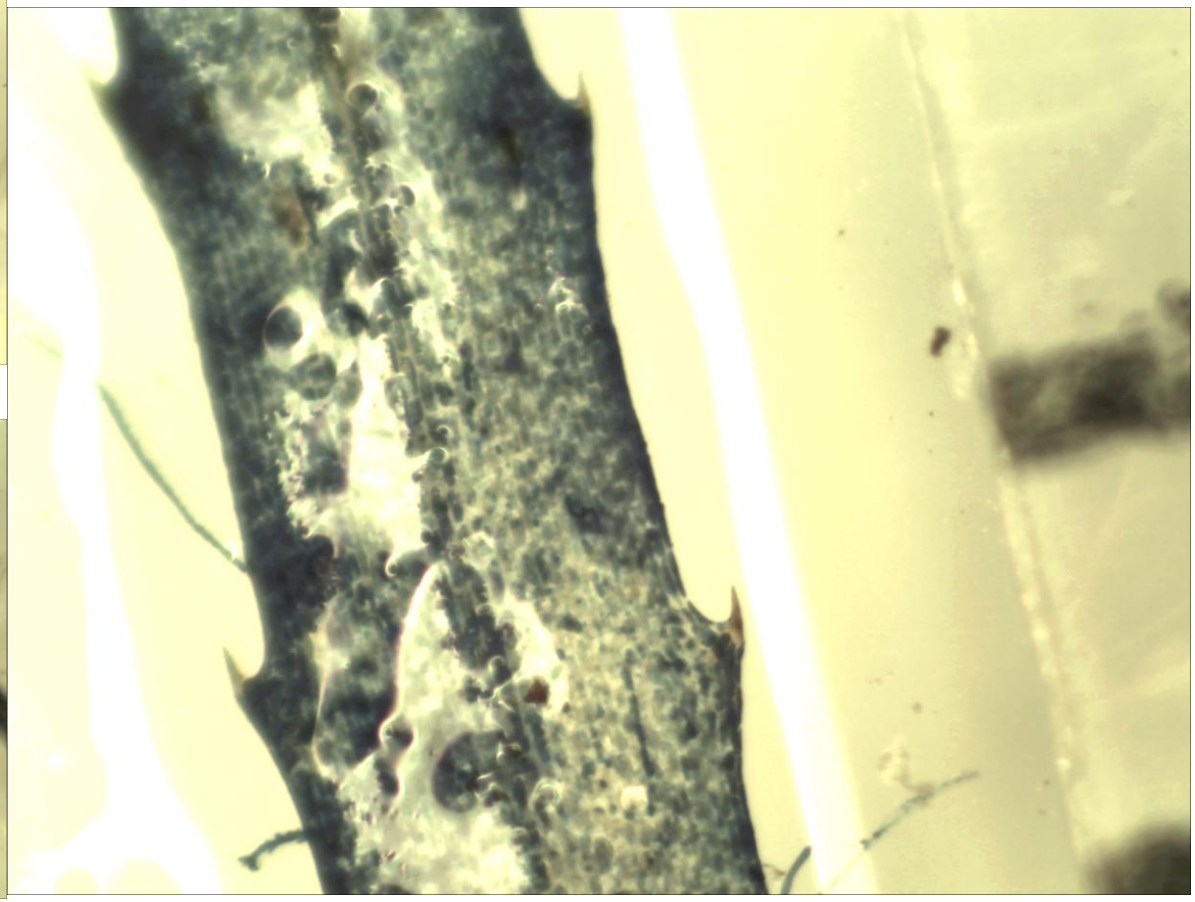




Najaflor006
G7

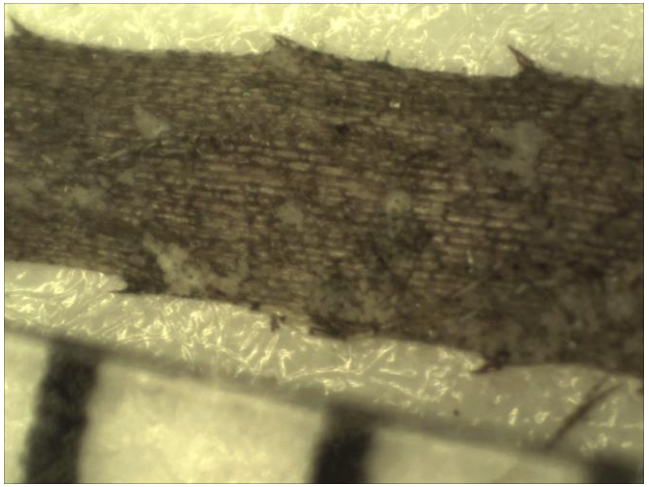


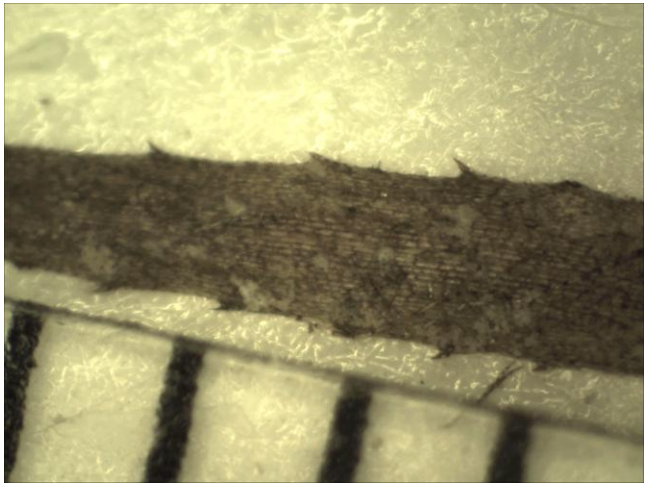




cG8
'floridana'
clade

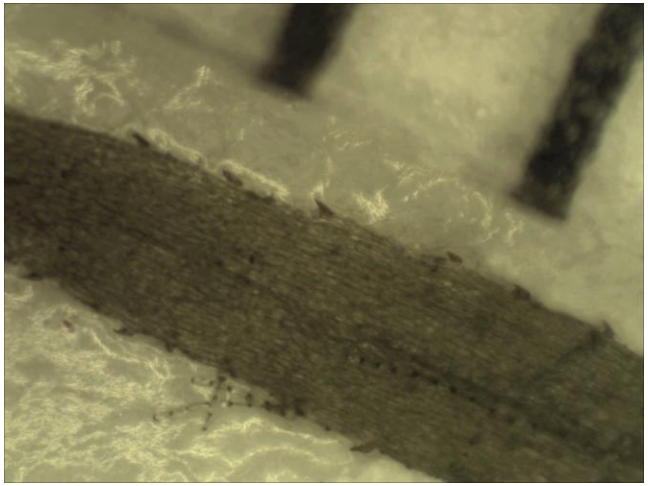
Najaguad010
But most teeth not exserted
FL

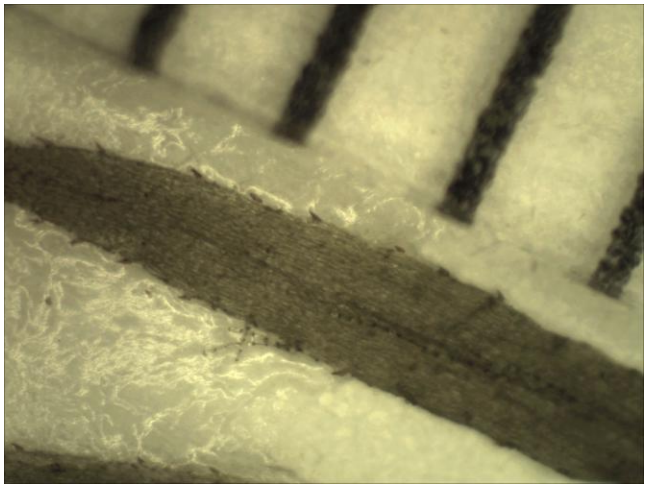




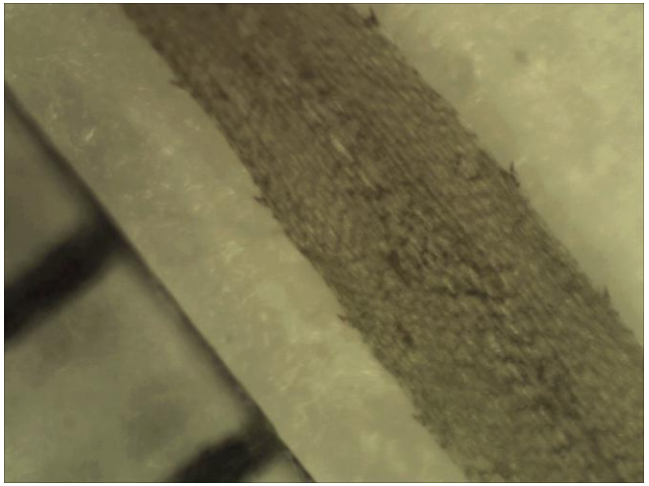
nrG1

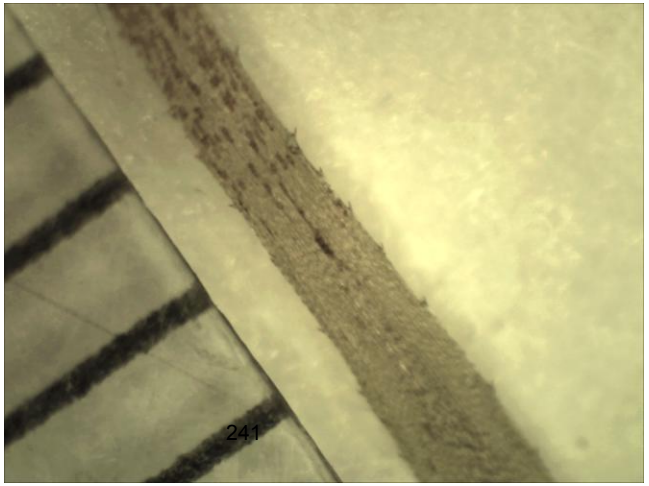
Najaguad235
cG10
WY (with introduced fish)





Najaguad017
cG10
AL



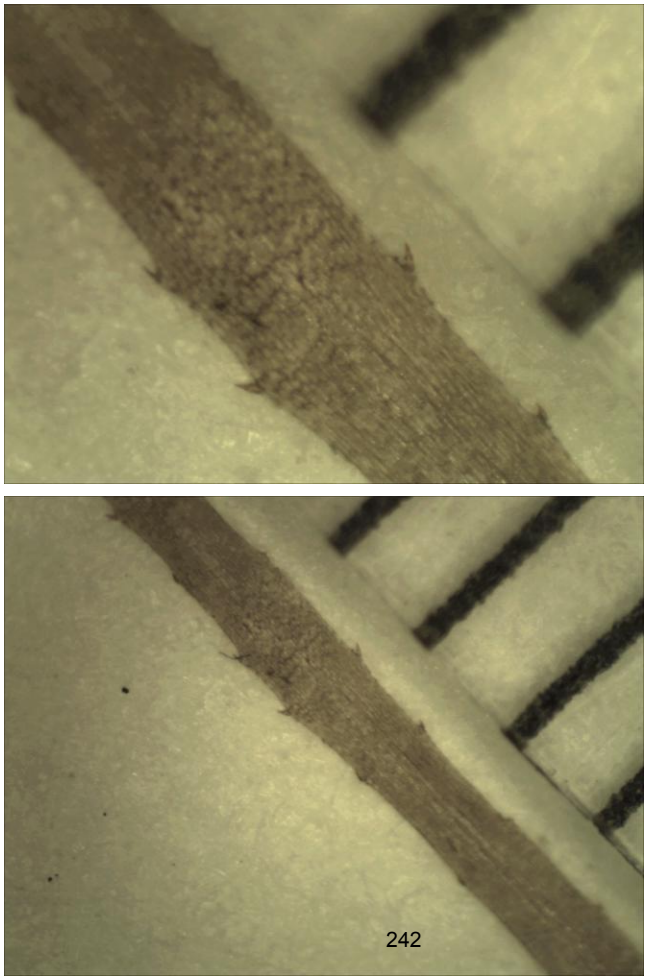


These accessions have the 'floridana' ribotype but widespread haplotype

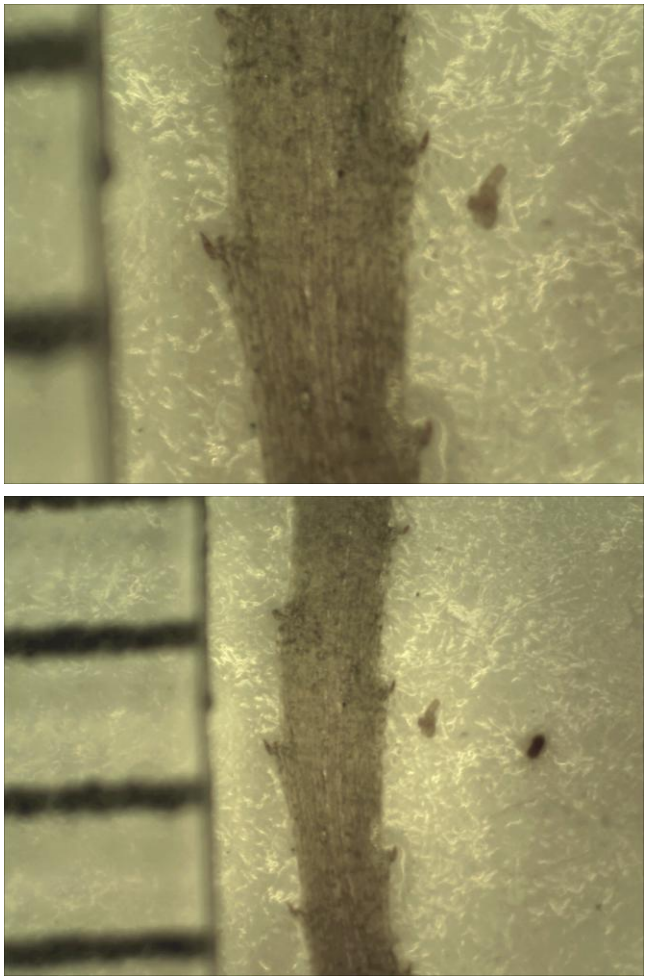
nrG1

These accessions have the 'floridana' ribotype but northern haplotypes

Najaguad015
cG16
SC



Najaflor001
cG16
nrG1 and nrG12
SC

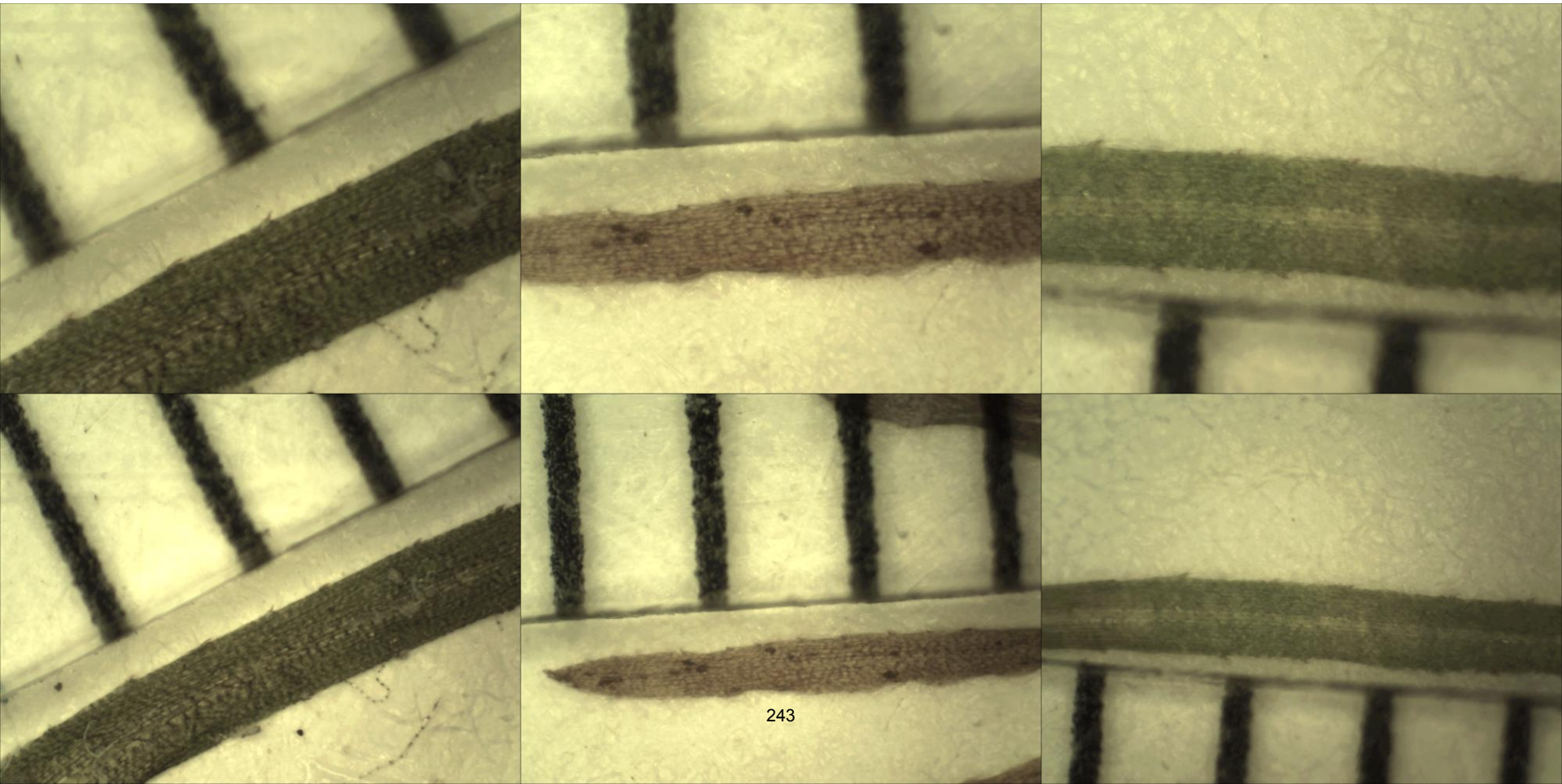


cG10

Najaguad061

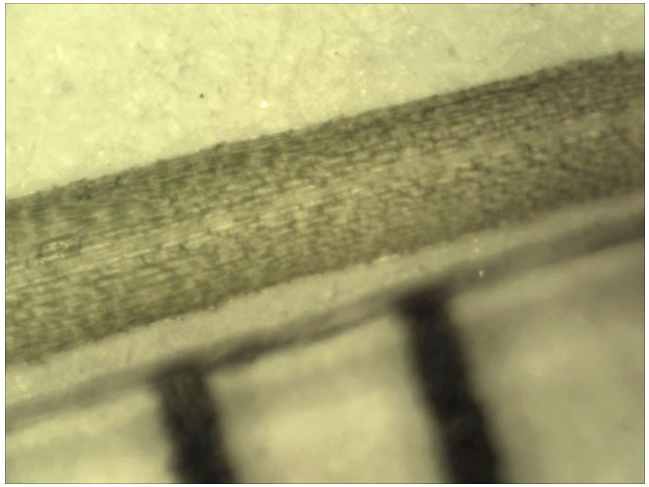
Najaguad032

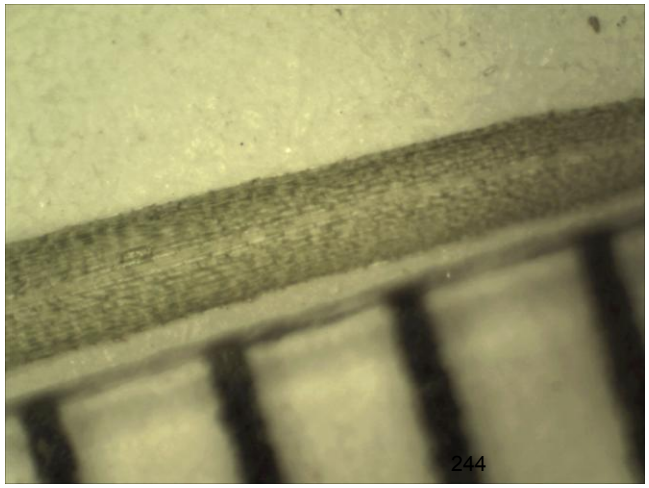
Najaguad018
GA



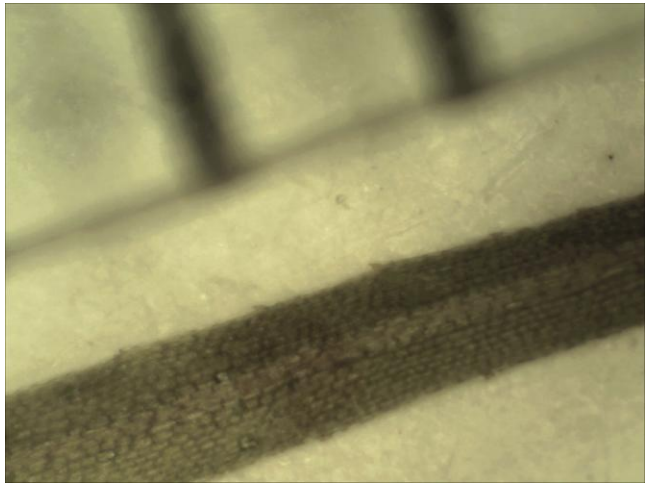
cG10

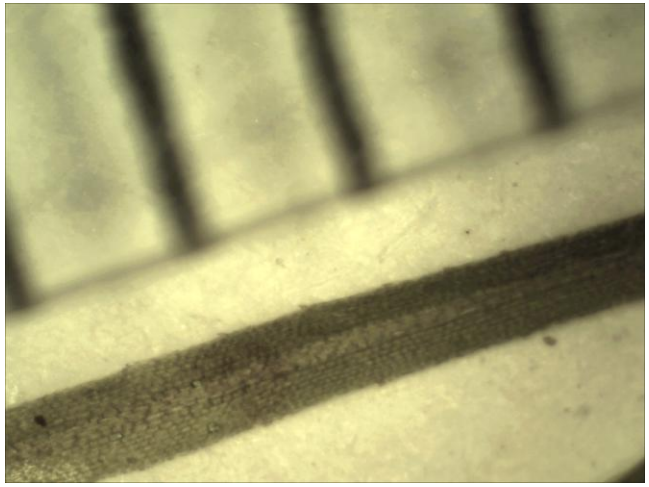
Najaguad172





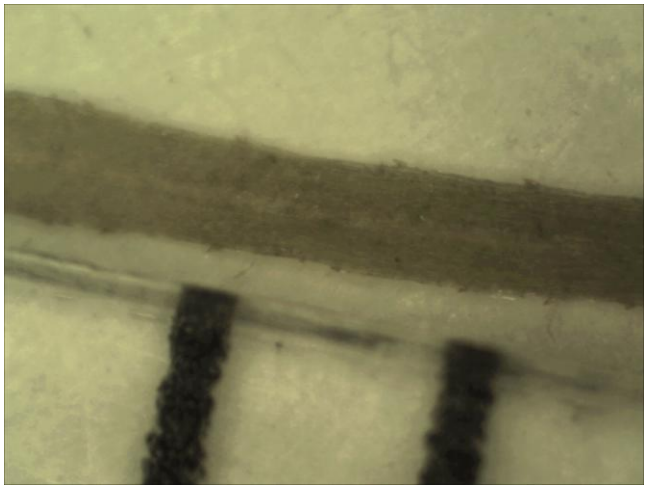
Najaguad076

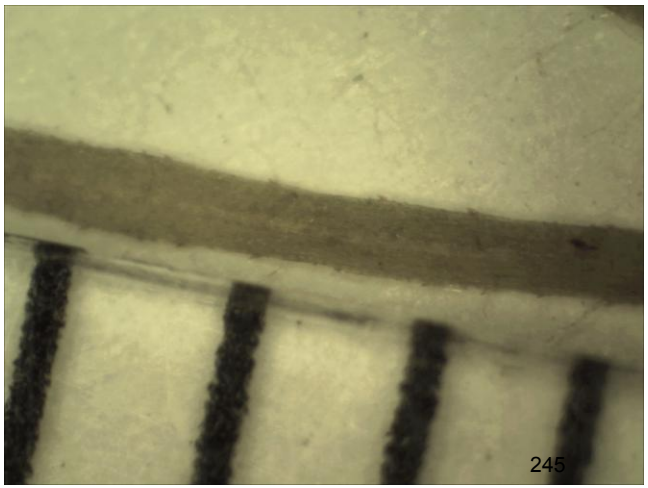




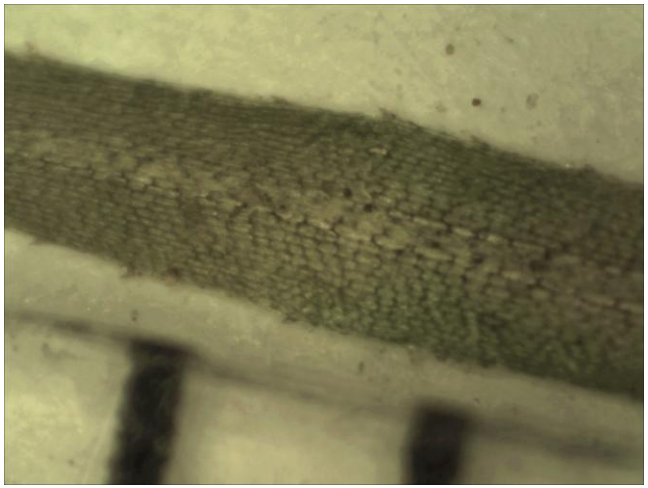
cG11

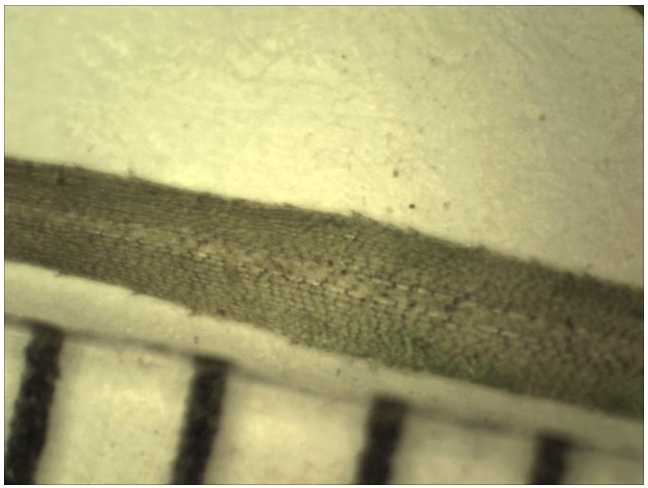
Najaguad063





Najaguad072



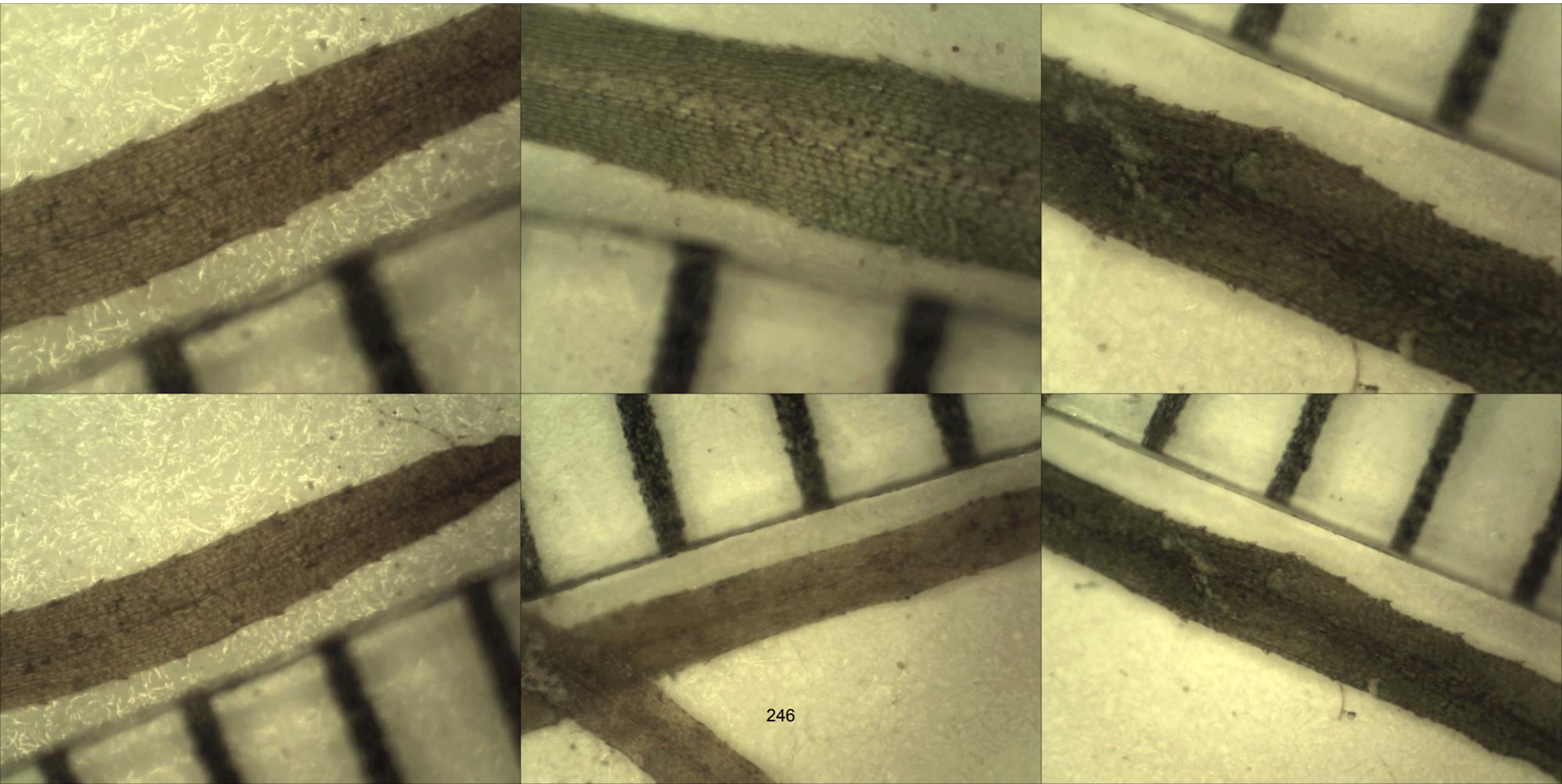


cG12

Najaguad031

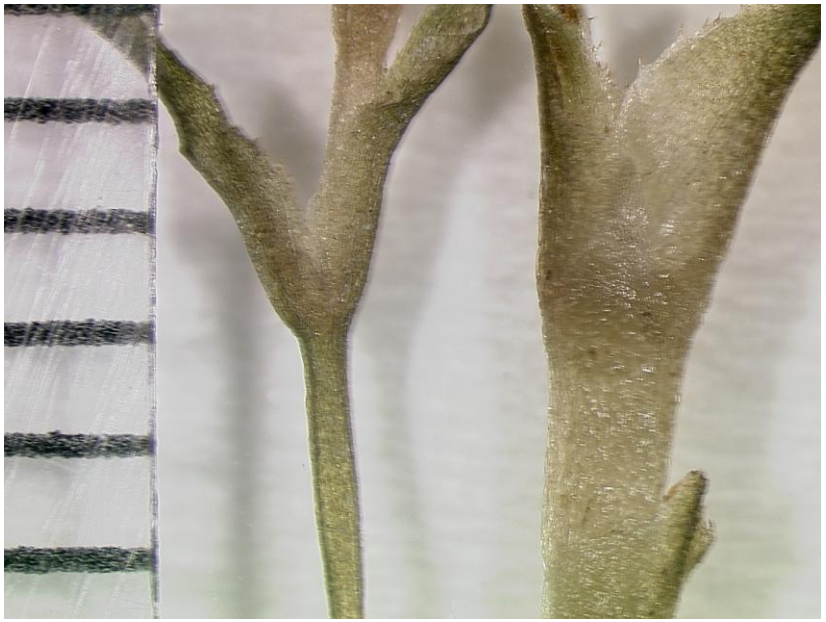
Najaguad071

Najaguad135



cG14 'olivacea' clade

Najaguad220 and Najaoliv253

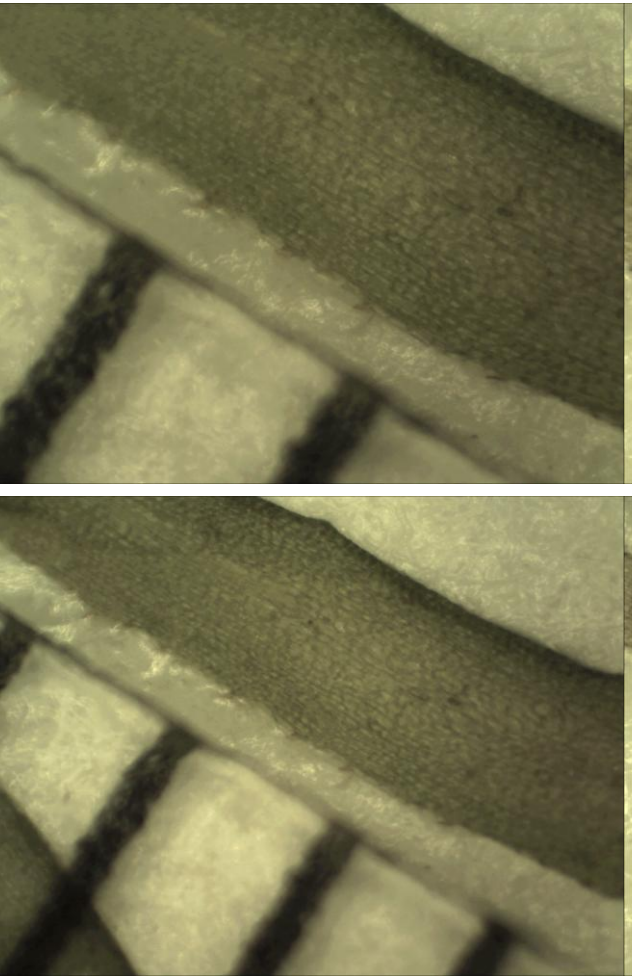
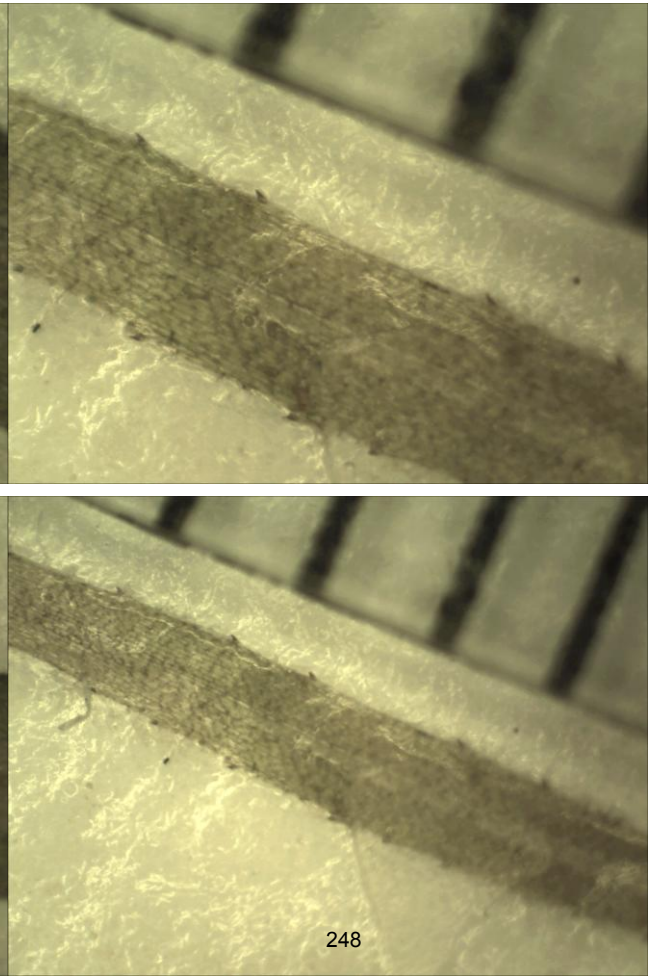
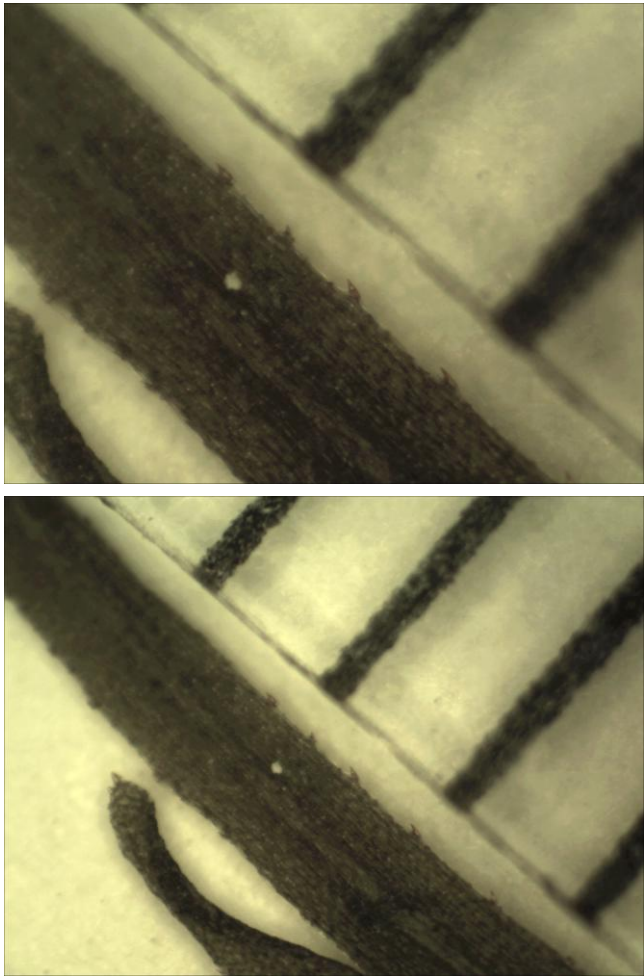


cG16

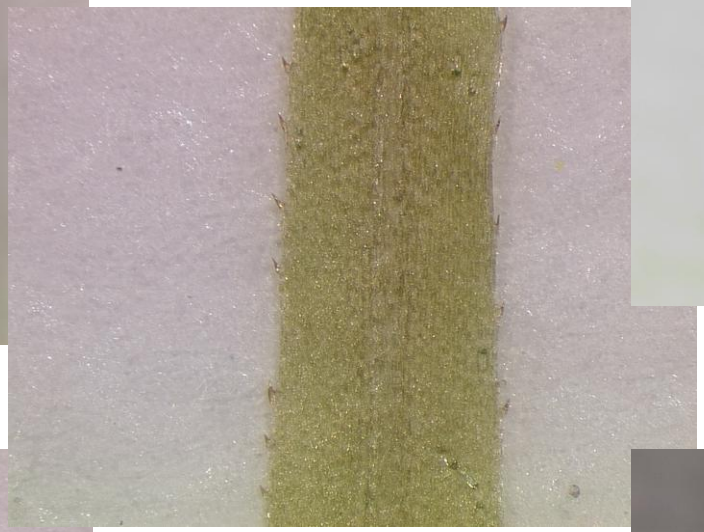
Najaguad114
OH

Najaguad111
IN

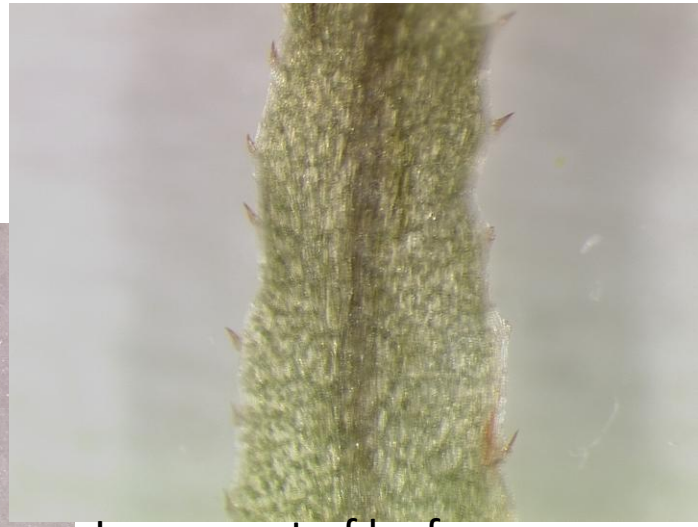
Najaguad097
MN

Najaflor006 and Najaguad062 and Najaoliv253 (57.5x)

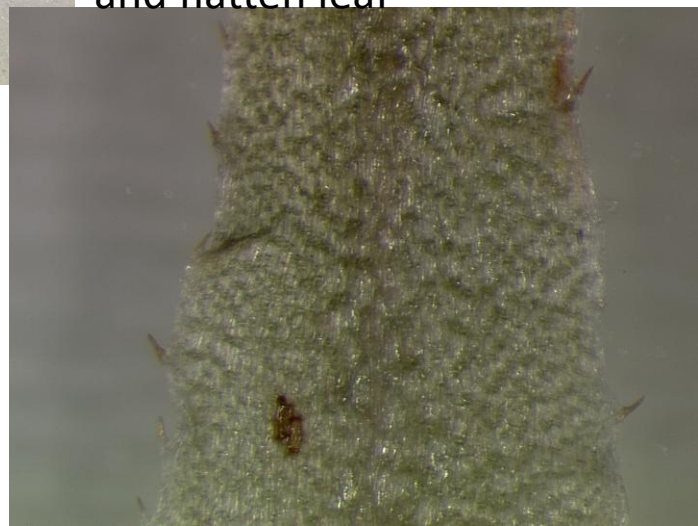


Mid leaf

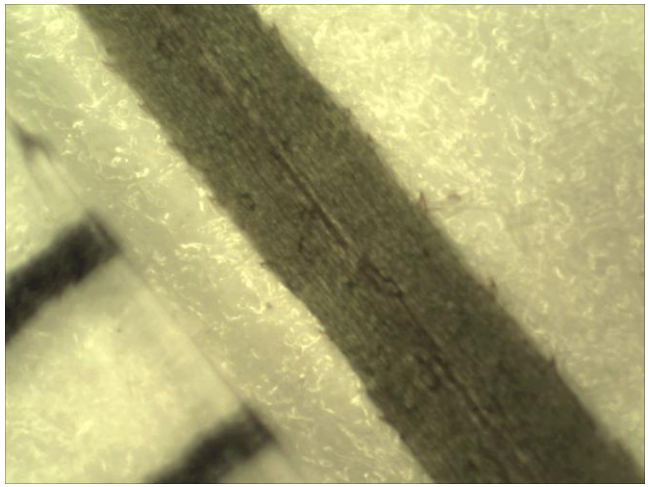


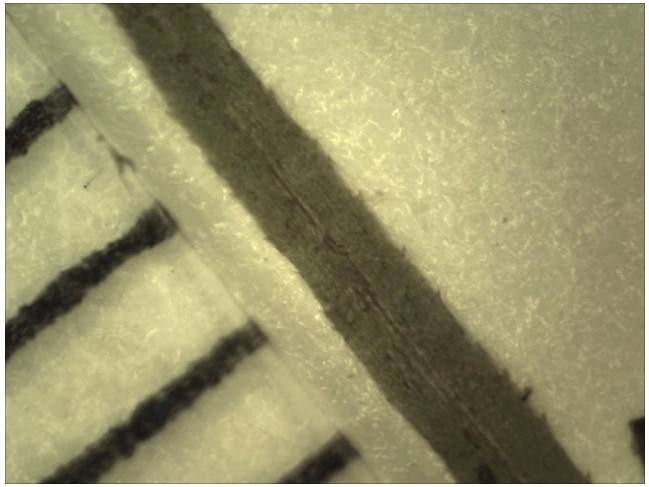
Lower part of leaf
Note: difficult to rehydrate and flatten leaf

1mm



Najaguad028
nrG9
rbFL1 – no matK
CO

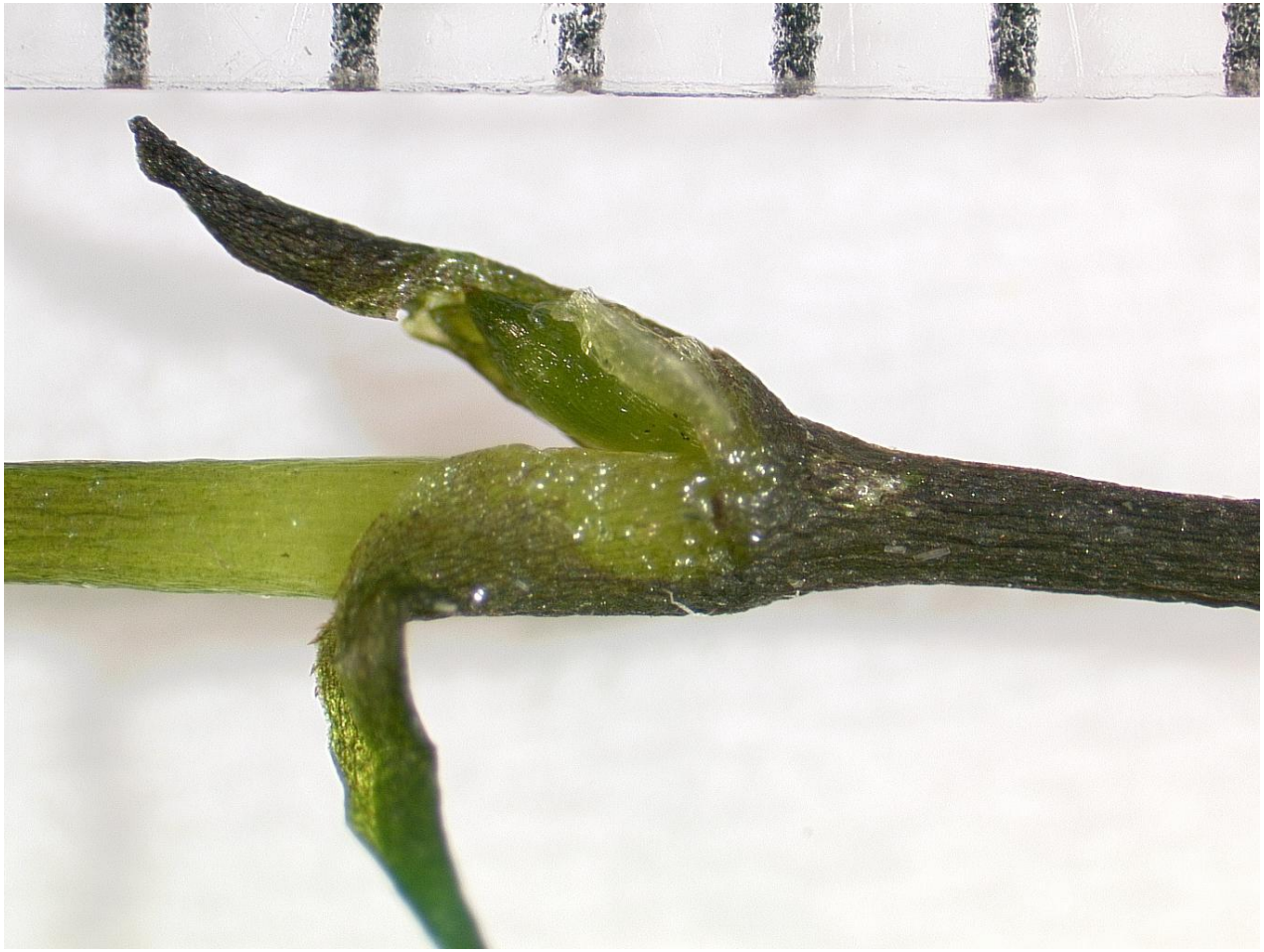




Appendix I
N. Guadalupensis
Mansfield Hollow,
Connecticut. 12th September

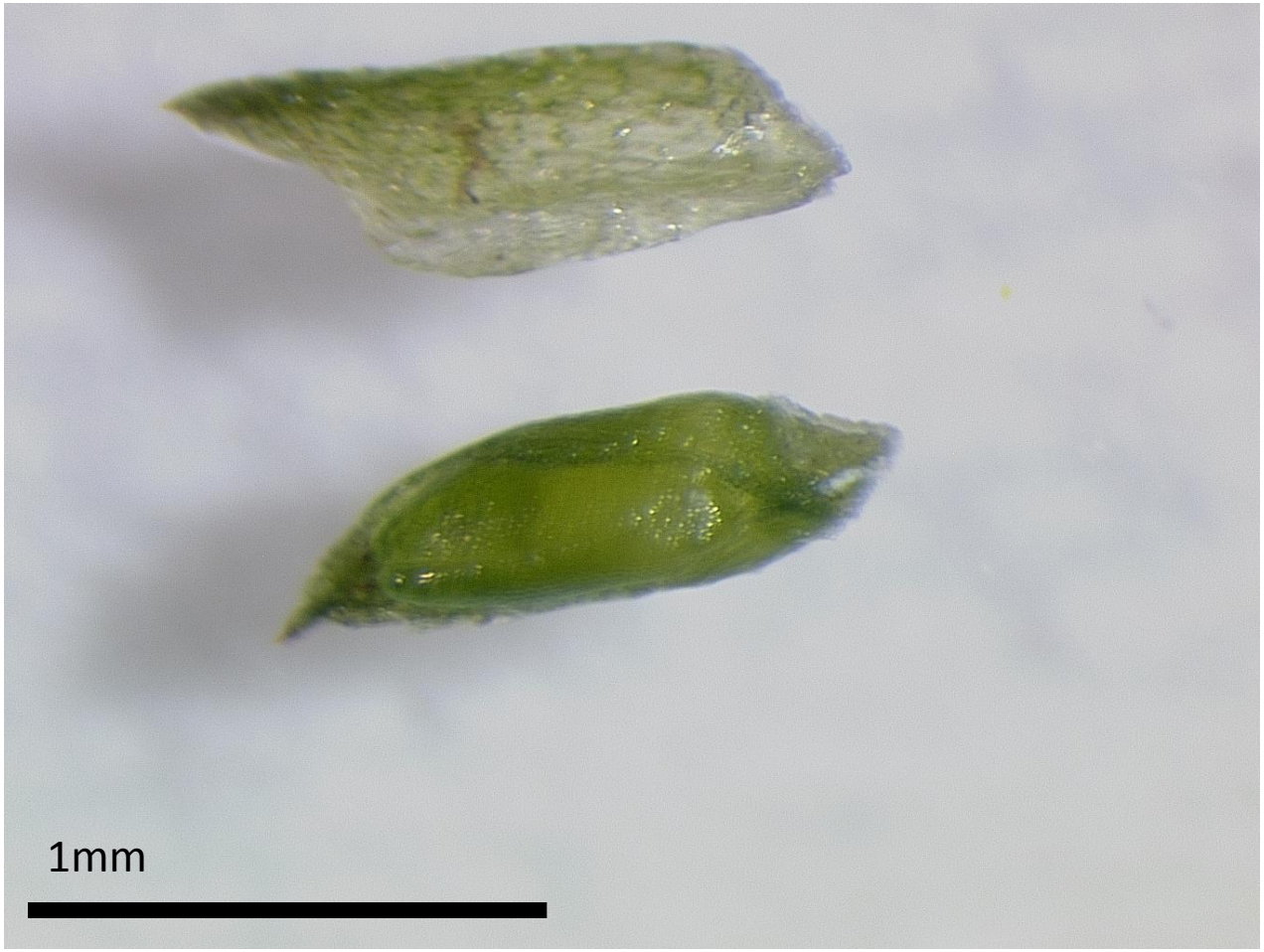

Node with turion

Node with turion

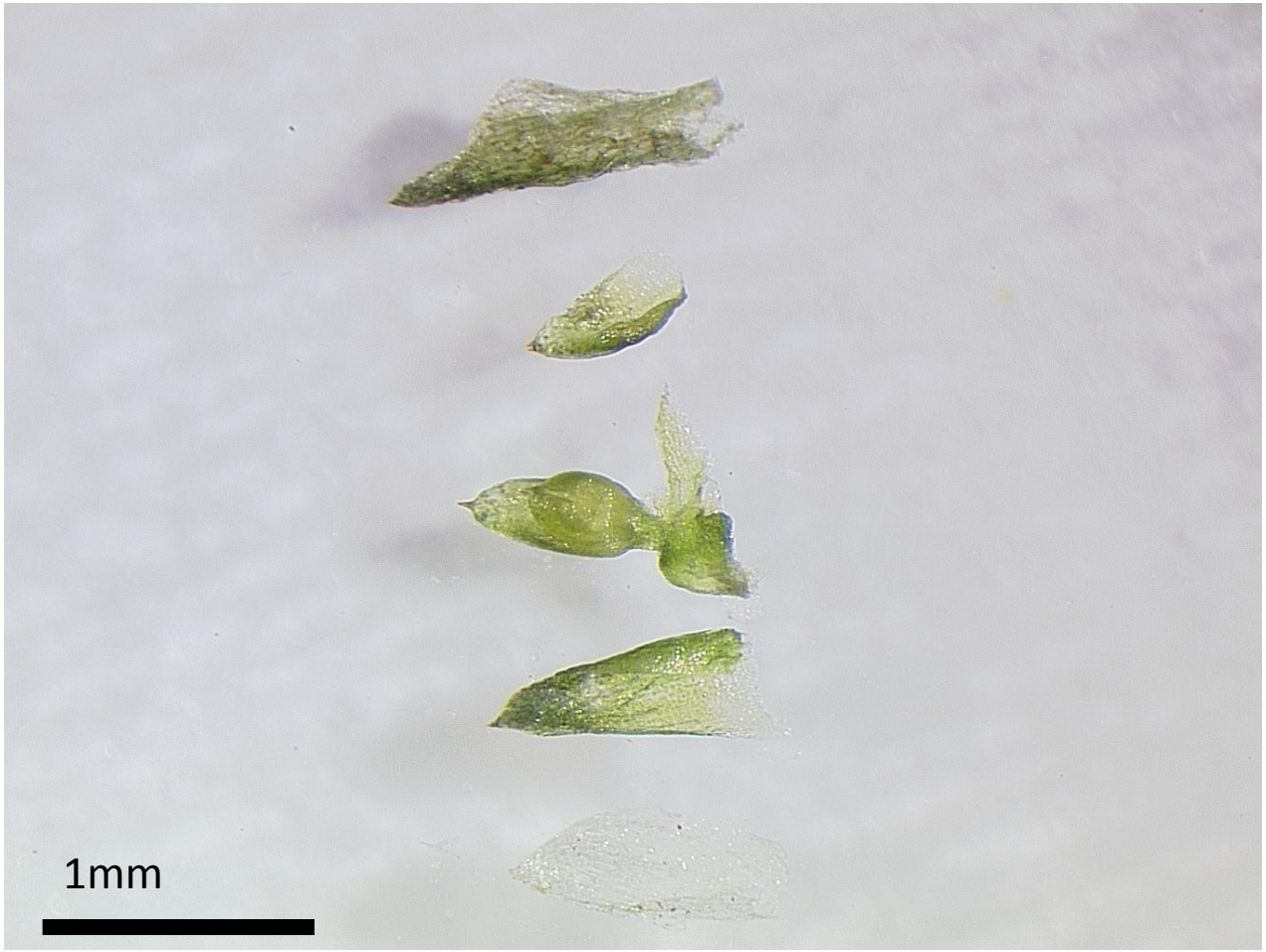


One leaf removed

Two leaves removed

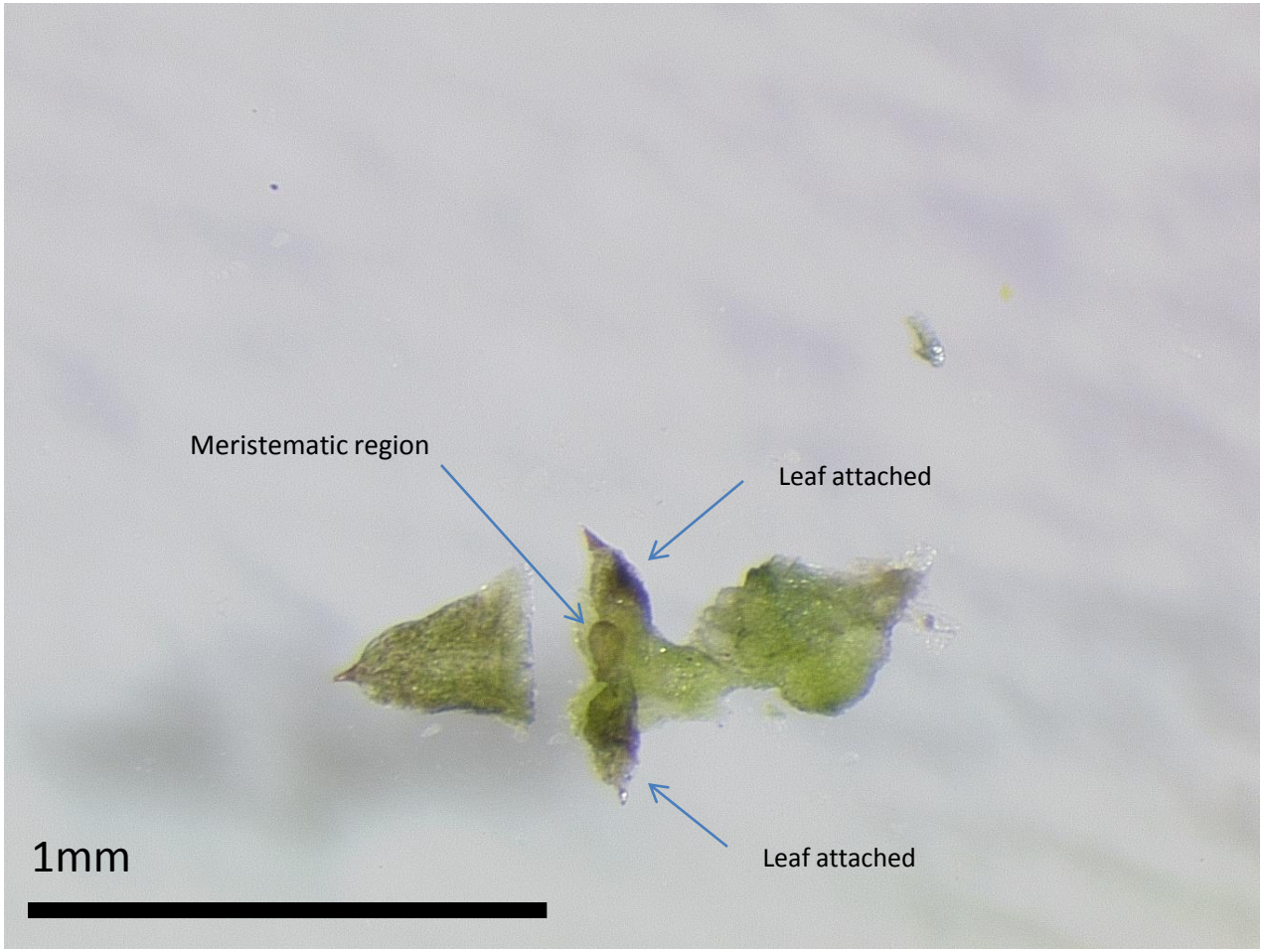


Five leaves removed



Five leaves removed

Six leaves removed (two still attached)



APPENDIX J

Taxa for Seed and flower counts

Lab number	Herbarium		Lab number	Herbarium		Lab number	Herbarium	
Najas flexilis (49)			**Najas canadensis (78)**			**Najas guadalupensis (177)**		
Najaflex008	CONN		Najacana009	CONN		Najaguad001	CONN	cG13
Najaflex012	CONN		Najacana011	CONN		Najaguad002	CONN	cG15
Najaflex017	CONN		Najacana015	CONN		Najaguad003	CONN	cG16
Najaflex018	CONN		Najacana019	CONN		Najaguad004	CONN	cG15
Najaflex027	CONN		Najacana020	CONN		Najaguad005	CONN	cG16
Najaflex035	CONN		Najacana028	CONN		Najaguad006	CONN	cG15
Najaflex042	CONN		Najacana031	CONN		Najaguad008	CONN	cG10
Najaflex043	CONN		Najacana033	CONN		Najaguad010	CONN	cG8
Najaflex044	CONN		Najacana036	CONN		Najaguad013	CONN	cG15
Najaflex048	CONN		Najacana040	CONN		Najaguad014	CONN	cG16
Najaflex049	CONN		Najacana041	CONN		Najaguad015	CONN	cG16
Najaflex050	CONN		Najacana047	CONN		Najaguad016	CONN	cG10
Najaflex054	CONN		Najacana051	CONN		Najaguad017	CONN	cG10
Najaflex056	CONN		Najacana052	CONN		Najaguad018	CONN	cG10
Najaflex066	CONN		Najacana053	CONN		Najaguad019	CONN	cG6
Najaflex068	CONN		Najacana057	CONN		Najaguad020	CONN	cG10
Najaflex069	CONN		Najacana060	CONN		Najaguad021	CONN	cG10
Najaflex070	CONN		Najacana061	CONN		Najaguad022	CONN	cG10
Najaflex072	CONN		Najacana062	CONN		Najaguad023	CONN	cG13
Najaflex074	CONN		Najacana064	CONN		Najaguad029	CONN	cG16
Najaflex080	CONN		Najacana067	CONN		Najaguad030	CONN	cG16
Najaflex081	CONN		Najacana075	CONN		Najaguad031	CONN	cG12
Najaflex084	CONN		Najacana077	CONN		Najaguad032	CONN	cG10
Najaflex085	CONN		Najacana078	CONN		Najaguad037	CONN	cG15
Najaflex086	CONN		Najacana083	CONN		Najaguad039	CONN	cG15
Najaflex087	CONN		Najacana089	CONN		Najaguad043	CONN	cG10
Najaflex088	CONN		Najacana091	CONN		Najaguad044	CONN	cG10
Najaflex090	CONN		Najacana092	CONN		Najaguad047	CONN	cG16
Najaflex094	CONN		Najacana093	CONN		Najaguad048	CONN	cG16

Najaflex096	CONN		Najacana103	CONN		Najaguad049	CONN	cG16
Najaflex098	CONN		Najacana104	CONN		Najaguad050	CONN	cG15
Najaflex099	CONN		Najacana107	CONN		Najaguad051	CONN	cG15
Najaflex100	CONN		Najacana114	CONN		Najaguad054	CONN	cG10
Najaflex101	CONN		Najacana115	CONN		Najaguad055	CONN	cG10
Najaflex102	CONN		Najacana116	CONN		Najaguad056	CONN	cG10
Najaflex109	CONN		Najacana117	CONN		Najaguad057	CONN	cG10
Najaflex110	CONN		Najacana118	CONN		Najaguad058	CONN	cG10
Najaflex111	CONN		Najacana119	CONN		Najaguad059	CONN	cG10
Najaflex122	CONN		Najacana120	CONN		Najaguad060	CONN	cG10
Najaflex129	CONN		Najacana121	CONN		Najaguad061	CONN	cG10
Najaflex1413	CONN		Najacana125	CONN		Najaguad062	CONN	cG3
Najaflex144	CONN		Najacana127	CONN		Najaguad063	CONN	cG11
Najaflex147	CONN		Najacana128	CONN		Najaguad064	CONN	cG10
Najaflex156	CONN		Najacana133	CONN		Najaguad065	CONN	cG3
Najaflex159	CONN		Najacana135	CONN		Najaguad066	CONN	cG10
Najaflex160	CONN		Najacana136	CONN		Najaguad067	CONN	cG10
Najaflex163	CONN		Najacana140	CONN		Najaguad068	CONN	cG10
Najaflex177	CONN		Najacana141	CONN		Najaguad069	CONN	cG10
Najaflex209	CONN		Najacana1411	CONN		Najaguad070	CONN	cG10
			Najacana142	CONN		Najaguad071	CONN	cG12
			Najacana143	CONN		Najaguad072	CONN	cG11
			Najacana146	CONN		Najaguad073	CONN	cG10
			Najacana154	CONN		Najaguad074	CONN	cG4
			Najacana157	CONN		Najaguad075	CONN	cG9
			Najacana158	CONN		Najaguad076	CONN	cG10
			Najacana161	CONN		Najaguad077	CONN	cG10
			Najacana179	CONN		Najaguad078	CONN	cG10
			Najacana182	CONN		Najaguad079	CONN	cG10
			Najacana187	CONN		Najaguad080	CONN	cG10
			Najacana188	CONN		Najaguad082	CONN	cG10
			Najacana195	CONN		Najaguad083	CONN	cG10
			Najacana196	CONN		Najaguad084	CONN	cG10
			Najacana200	CONN		Najaguad085	CONN	cG10

			Najacana201	CONN		Najaguad086	CONN	cG10
			Najacana204	CONN		Najaguad087	CONN	cG10
			Najacana205	CONN		Najaguad088	CONN	cG10
			Najacana206	CONN		Najaguad089	CONN	cG10
			Najacana224	CONN		Najaguad090	CONN	cG10
			Najacana225	CONN		Najaguad091	CONN	cG10
			Najacana226	CONN		Najaguad092	CONN	cG10
			Najacana227	CONN		Najaguad093	CONN	cG12
			Najacana228	CONN		Najaguad094	CONN	cG2
			Najacana229	CONN		Najaguad095	CONN	cG2
			Najacana230	CONN		Najaguad096	CONN	cG10
			Najacana231	CONN		Najaguad097	CONN	cG16
			Najacana232	CONN		Najaguad099	CONN	cG16
			Najacana239	CONN		Najaguad100	CONN	cG16
			Najacana248	CONN		Najaguad101	CONN	cG16
						Najaguad102	CONN	cG16
						Najaguad103	CONN	cG16
						Najaguad104	CONN	cG16
						Najaguad105	CONN	cG16
						Najaguad106	CONN	cG16
						Najaguad107	CONN	cG16
						Najaguad108	CONN	cG16
						Najaguad111	CONN	cG16
						Najaguad112	CONN	cG16
						Najaguad113	CONN	cG16
						Najaguad114	CONN	cG16
						Najaguad117	CONN	cG16
						Najaguad118	CONN	cG16
						Najaguad119	CONN	cG16
						Najaguad121	CONN	cG16
						Najaguad122	CONN	cG15
						Najaguad123	CONN	cG16
						Najaguad124	CONN	cG3
						Najaguad125	CONN	cG16

						Najaguad132	CONN	cG5
						Najaguad133	CONN	cG5
						Najaguad135	CONN	cG12
						Najaguad136	CONN	cG16
						Najaguad137	CONN	cG15
						Najaguad138	CONN	cG16
						Najaguad142	CONN	cG15
						Najaguad150	CONN	cG10
						Najaguad152	CONN	cG16
						Najaguad153	WI NBIC#652200	cG16
						Najaguad154	WI NBIC#652200	cG16
						Najaguad155	CONN	cG16
						Najaguad156	CONN	cG16
						Najaguad157	WI NBIC#677400	cG16
						Najaguad160	CONN	cG16
						Najaguad161	CONN	cG14
						Najaguad163	CONN	cG16
						Najaguad164	CONN	cG16
						Najaguad168	CONN	cG10
						Najaguad169	CONN	cG13
						Najaguad170	CONN	cG12
						Najaguad171	CONN	cG10
						Najaguad172	CONN	cG10
						Najaguad173	CONN	cG10
						Najaguad174	CONN	cG10
						Najaguad175	CONN	cG12
						Najaguad176	CONN	cG12
						Najaguad177	CONN	cG3
						Najaguad178	CONN	cG12
						Najaguad179	CONN	cG16
						Najaguad181	CONN	cG16
						Najaguad183	CONN	cG16
						Najaguad184	CONN	cG16
						Najaguad185	CONN	cG16

						Najaguad186	CONN	cG15
						Najaguad192	CONN	cG16
						Najaguad193	CONN	cG16
						Najaguad194	CONN	cG15
						Najaguad195	CONN	cG16
						Najaguad196	CONN	cG2
						Najaguad206	CONN	cG15
						Najaguad212	CDA	cG13
						Najaguad214	CDA	cG13
						Najaguad216	CONN	cG16
						Najaguad218	CONN	cG16
						Najaguad219	CONN	cG14
						Najaguad220	CONN	cG14
						Najaguad221	CONN	cG16
						Najaguad223	CONN	cG10
						Najaguad224	CONN	cG16
						Najaguad225	CONN	cG16
						Najaguad226	CONN	cG16
						Najaguad228	CONN	cG10
						Najaguad229	CONN	cG10
						Najaguad233	CONN	cG16
						Najaguad235	CONN	cG10
						Najaguad236	CONN	cG10
						Najaguad237	CONN	cG10
						Najaguad243	CONN	cG16
						Najaguad244	CONN	cG2
						Najaguad245	CONN	cG10
						Najaguad247	CONN	cG14
						Najaguad248	CONN	cG15
						Najaguad251	CONN	cG15
						Najaguad255	CONN	cG1
						Najaguad256	CONN	cG16
						Najaflor001	CONN	cG16
						Najaflor002	CONN	cG7

						Najaflor003	CONN	cG7
						Najaflor006	CONN	cG7
						Najaflor007	CONN	cG7
						Najaflor009	CONN	cG7
						Najaflor010	CONN	cG7
						Najaflor014	CONN	cG7
						Najaflor015	CONN	cG7
						Najaflor016	CONN	cG7
						Najaoliv098	CONN	cG14
						Najaoliv238	MICH 1485091	cG14
						Najaoliv241	CONN	cG14
						Najaoliv253	CONN	cG14

Chapter 3 List of appendices

Appendix A. Taxon sampling, Illumina/454 reads, chloroplast characteristics, and repeat analysis.

Appendix B. ORFs. and ndh genes

Appendix C. Plastome patterns in 106 angiosperms

APPENDIX A

Taxon sampling

Lab number	Taxon	Country	State	County	Lake	Collector	Latitude	Longitude	Roche 454 reads	Trimmed	Illumina Paired reads	Trimmed paired	Assembled
Najamari022	N. major All.	Czech Republic	Bohemia	Hradec Králové	Lake in sand pit, 1 km	Zdenek Kaplan 10/316	50.159444	15.541388	219676	207886	1518846	1237040	31116
Najamari016	N. marina L.	USA	WI	Jefferson County	Lake Ripley	D.H. Les 970	43.004519	-88.988349	328886	311576	1358822	1084756	27614
Najamino057	N. minor (USA1)	USA	PA	Mercer County	Wilhelm Lake	R.K. Shannon 1251	41.423066	-80.149874	81765	77204	5047208	4037866	154960
Najamino061	N. minor (USA2)	USA	NY	Putnam County	China Pond	A. Les 11-0823	41.440282	-73.740369	10244	9656	8270268	6618215	225012
Najagrac017	N. gracillima	USA	MN	Hubbard County	Wabigish Lake	D.H. Les 931	47.116666	-95.118055	29439	27624	886916	860543	96767
Najafili001	N. filifolia	USA	GA	Decatur County	Pond near Bainbridge	D. H. Les and N. P. Tippery 756/219	30.955680	-84.436413			894974	753956	74428
Najawrig002	N. wrightiana	USA	FL	Collier County	Big Cypress Swamp, E	W.T. Haller s/n	25.985501	-81.056585			380804	373780	24081
Najacana206	N. canadensis	USA	NY	Rensselaer County	Troy Reservoir	D.H. Les 1123	42.764903	-73.633750			8313046	6680872	136053
Najaflex100	N. flexilis	USA	CT	Litchfield County	Twin Lakes	R.K. Shannon 1157	42.019506	-73.389567			7414420	6438070	133687
Hydrilla064	Hydrilla verticillata	USA	FL	Okeechobee County	Lake Okeechobee	L.K. Benoit	26.968889	-80.797500	272706		3025132	2353518	67639
Note:	N. major All. [= N. marina subsp. marina = Karyotype A (Viinikka 1976) = ITSA (Rüegg et al. 2016)]												
	N. marina L. [= N. marina subsp. intermedia = Karyotype B (Viinikka 1976) = ITSB (Rüegg et al. 2016)]												

Chloroplast characteristics

	Najamari022	Najamari016	Najamino057	Najamino061	Najagrac017	Najafili001	Najawrig002	Najacana206	Najaflex100
Size (bp)	159886	158009	156821	156791	158592	160876	161478	161390	161093
LSC length (bp)	88366	88543	87872	87842	87427	88321	89307	88987	89653
IR length (bp)	34359	33394	33179	33178	33993	34614	34390	34526	34606
SSC length (bp)	2802	2678	2591	2593	3179	3327	3391	3351	2228
Overall % GC content (%) *	37.9	37.9	38.2	38.2	38.4	38.1	38.1	37.9	38
LSC % GC content	35.2	35.1	35.6	35.6	35.9	35.6	35.5	35.3	35.2
SSC % GC content	27.3	28.4	28.1	28	29.3	28	27.9	27.2	28.1
IR % GC content	41.8	41.9	42	42	42	41.9	41.9	41.9	41.8
% Protein coding *	39.1	39	39.1	39.1	39.2	39.2	39.1	39	39.1
% rRNA *	54.7	54.7	54.8	54.8	54.8	54.9	54.9	54.9	54.8
% tRNA *	52.3	52.2	52.3	52.3	52.3	52.3	52.3	52.2	52.3
* one IR removed									

	NC_018541	NC_029815	NC_029814	DQ400350	NC_015899	NC_015894	NC_029813	NC_016753	NC_007407	NC_010093	Z00044
	Elodea canadensis	Sagittaria lichuanensis	Potamogeton perfoliatus	Lemna minor	Wolffiella australiana	Wolffiella lingulata	Tofieldia thibetica	Colocasia esculenta	Acorus calamus	Acorus americanus	Nicotiana tabaccum
Size (bp)	156700	179007	156226	165955	168704	169337	155512	162424	153821	153819	155943
LSC length (bp)	86194	99125	86764	89907	91454	92015	86194	89670	84149	83496	86686
IR length (bp)	26349	33302	25612	31222	31930	31683	26389	25273	25697	26025	25343
SSC length (bp)	17808	13278	18238	13604	13392	13956	18150	22208	18278	18273	18571
Overall % GC content (%) *	37	36.8	36.5	35.7	35.9	35.8	37.4	36.2	38.6	38.6	37.8
LSC % GC content	34.8	34.7	34.2	33.5	33.8	33.7	33.3	34.4	37.2	37.2	35.9
SSC % GC content	30.5	30.6	29.6	30.3	30.8	30.8	31.6	28.9	33.4	33.4	32.1
IR % GC content	42.7	42.7	42.7	40.1	39.9	40	42.8	42.4	42.7	42.6	43.2
% Protein coding *	37.7	38.9	37.4	37.4		37.3	37.4	37.9	37.9	38.9	38.9
% rRNA *	55.3	54.8	54.9	54.8		54.6	54.7	55.2	55	55.3	55.3
% tRNA *	52.8	53.2	52.1	52.4		52.2	52.7	52.7	51.5	53.3	53.2
* one IR removed											

Repeat analysis

STRs (min no. of repeat units)	Najamari022	Najamari016	Najamino057	Najamino061	Najagrac017	Najafili001	Najawrig002	Najacana206	Najaflex100
Mononucleotide >/= 8 bp	91	81	88	86	74	76	82	73	72
di- (5)	16	21	15	14	13	14	14	19	17
tri- (4)	1	1	1	2	1	3	2	3	3
tetra- (3)	9	9	5	5	9	11	9	9	8
penta- (3)	0	0	2	2	0	4	2	1	2
hexa-(3)	1	1	0	0	0	1	0	0	2

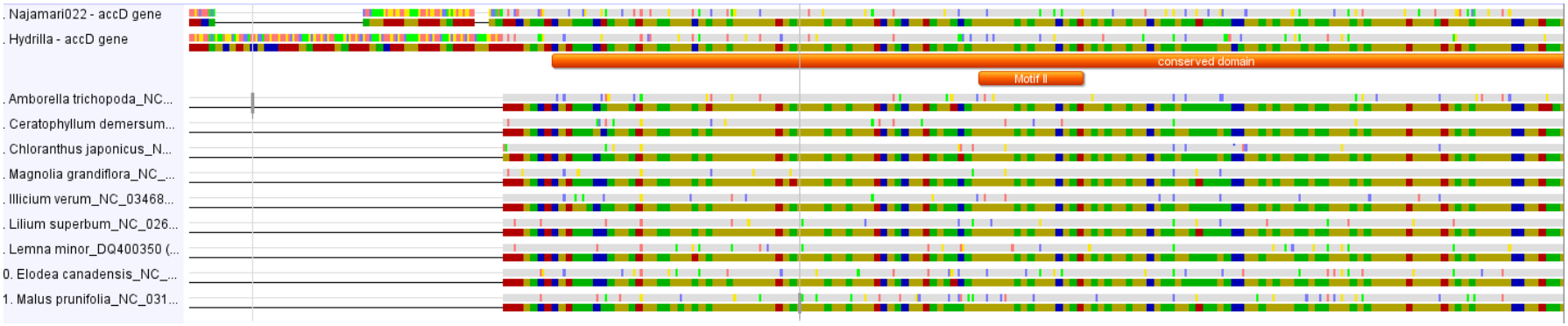
STRs (min no. of repeat units)	Elodea canadensis	Sagittaria lichuanensis	Potamogeton perfoliatus	Lemna minor	Wolffiella australiana	Wolffiella lingulata	Tofieldia thibetica	Colocasia esculenta	Acorus calamus	Acorus americanus	Nicotiana tabaccum
Mononucleotide >/= 8 bp	87	107	151	127	135	141	158	181	114	113	110
di- (5)	14	28	21	16	11	17	17	32	6	6	17
tri- (4)	5	15	3	6	5	5	3	9	5	5	3
tetra- (3)	12	10	15	10	15	11	13	15	4	4	13
penta- (3)	2	2	0	1	0	3	4	1	3	3	4
hexa-(3)	1	9	0	0	0	0	0	0	1	1	0

APPENDIX B

accD - coding for the plastid encoded subunit of Acetyl-CoA carboxylase, which regulates the rate of de novo fatty acid biosynthesis.

Alignable portion of the accD orf at the 3' end of gene. Both Najas and Hydrilla contain the 5 conserved motifs in the C-terminal region, including motif II, the suggested catalytic site (Lee et al. 2004).

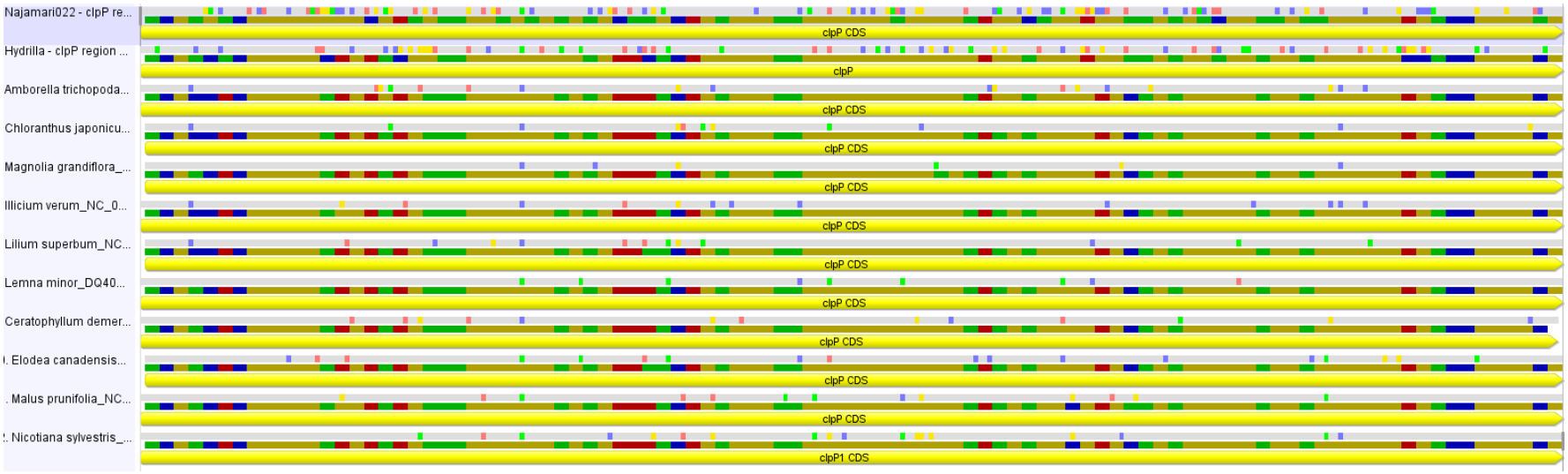
Repetitive motifs are associated with the gene region where Najas and Hydrilla diverge from other angiosperms.



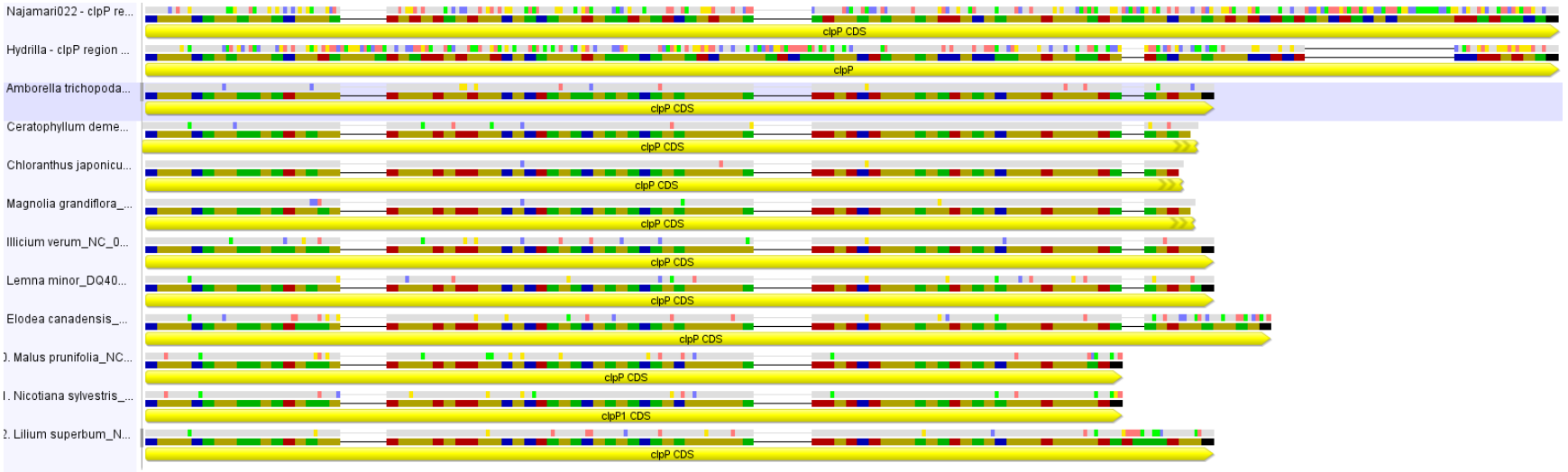
clpP1 - a proteolytic subunit of the ATP-ase dependent protease

clpP1-exon 1 and intergenic region, showing 30 bp insertion in Najas relative to other angiosperms here, and divergence of conserved monocot NEP -53 Type 1 promoter in Najas and Hydrilla.

clpP1 exon 2



clpP1 exon 3

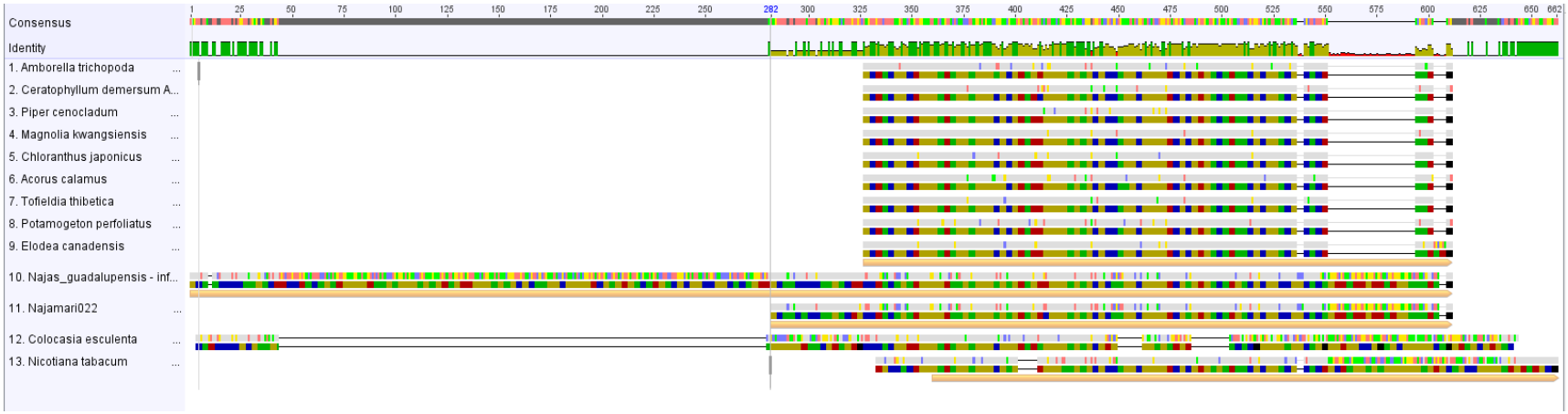


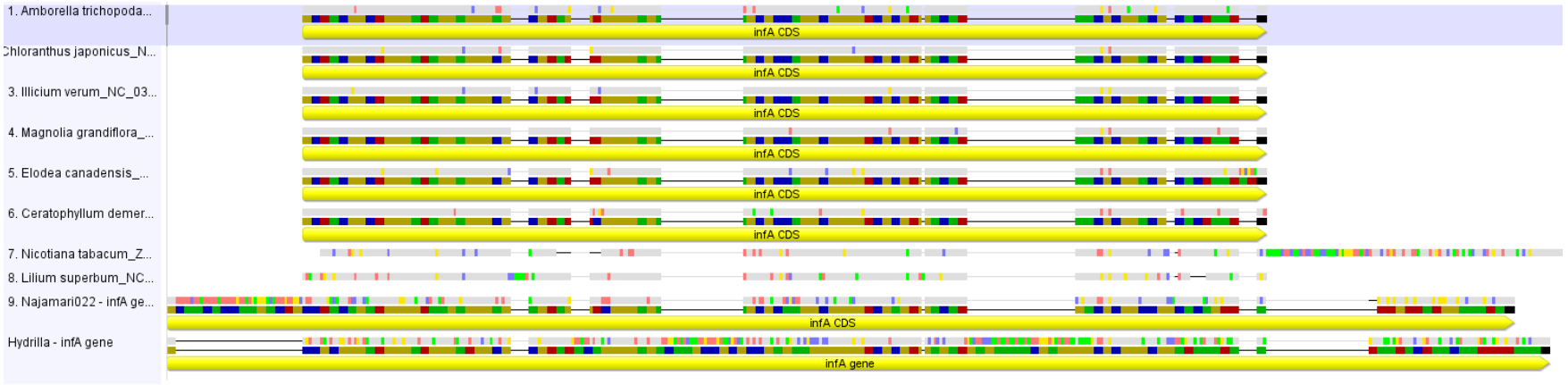
infA gene coding for translation initiation factor 1.

Comparision of the infA orf in N. major (Najamari022) seqenced in this study and N. guadalupensis (Ross et al.) with other angiosperms

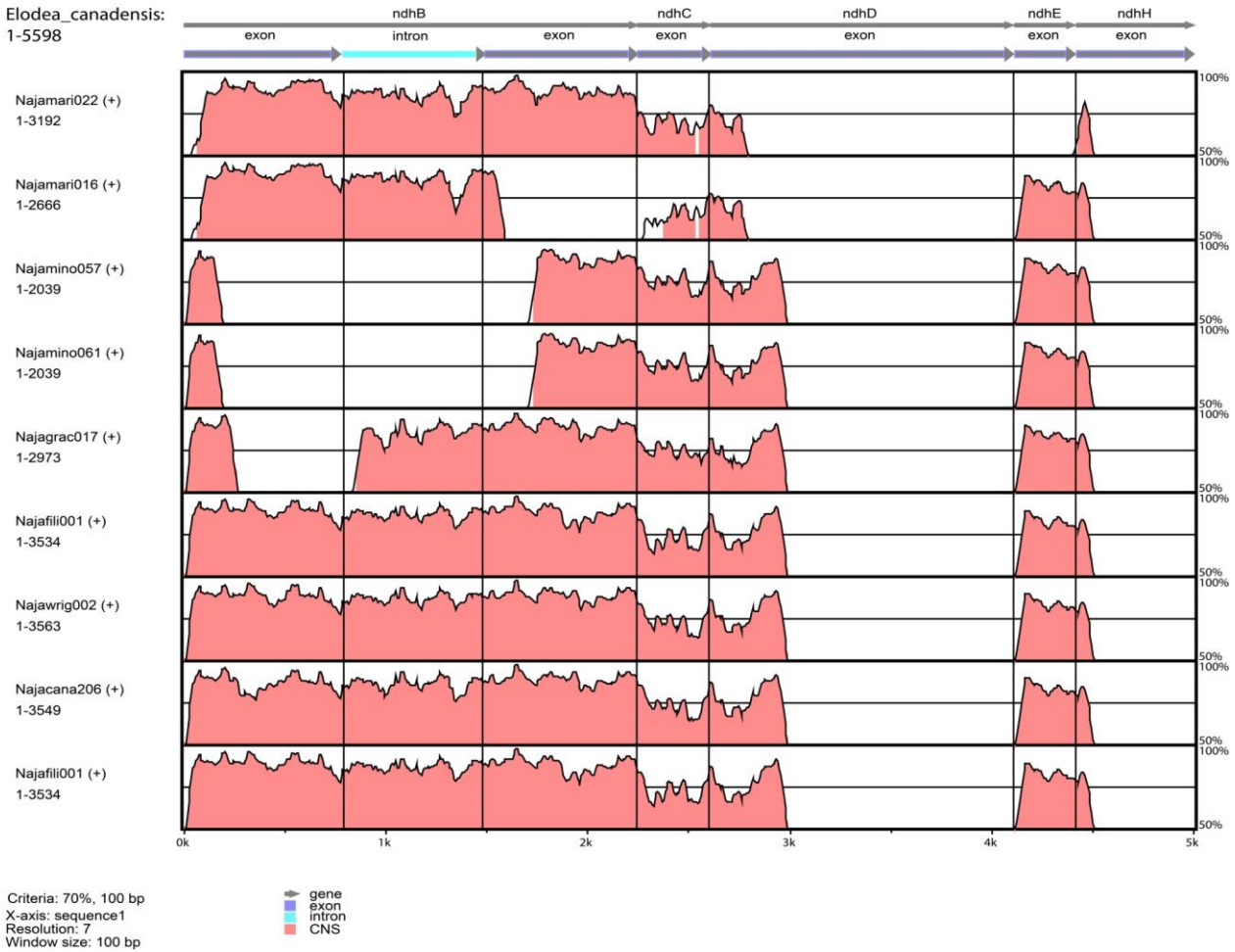
No sequence similarity found for this gene was predicted in DOGMA and No blast hit with megablast.

infA is a pseudogene in Colocasia and Nicotiana





mVISTA alignment of the five ndh pseudogenes in Najas, compared with Elodea



APPENDIX C

Regular plastomes

APPENDIX C

*** gene incorrectly annotated as having the intron missing
Distance = Distance from -10 to 16s RNA
accD/CM = are 5 conserved motifs of Lee et al. 2004 present?
CT = are the 3 amino acid residues of the the catalytic triad present (Wang et al. 1997)
1 IR=only one inverted repeat

	GenBank	Species	Family	Distance	RUA	minus 35	minus 10	rpo genes	ndh genes	accD	accD/CM	accD direction	clpP1	clpP1 direction	CT	infA (234)	rpl16 intron	IR perturbations
1	NC_034985	Acacia dealbata	Fabaceae	145	yes	yes	yes				yes	for	2 copies in LSC	no introns	yes	gone	intron	
2	NC_026134	Acacia ligulata	Fabaceae	145	yes	yes	yes				yes	for		2 introns-rev	no	gone	intron	
3	NC_034346	Acer griseum	Sapindaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	195	intron	
4	NC_010093	Acorus americanus	Acoraceae	145	yes	yes	yes				gene gone	n/a		2 introns-rev		234	intron	
5	NC_007407	Acorus calamus	Acoraceae	145	yes	yes	yes				gene gone	n/a		2 introns-rev		234	intron	
6	NC_032053	Agave americana	Asparagaceae	143	yes	yes	yes				yes	for		2 introns-rev		234	intron***	
7	NC_031829	Allium sativum	Amaryllidaceae	143	yes	yes	yes				yes	for		2 introns-rev		215	intron	
8	NC_005086	Amborella trichopoda	Amborellaceae	145	yes	yes	yes				no-5(1*)	for		2 introns-rev	yes	234	intron	
9	NC_015113	Anthriscus cerefolium	Apiaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	234	intron	
10	NC_000932	Arabidopsis thaliana	Brassicaceae	145	yes	yes	yes				yes	for		2 introns-rev		gone	intron	
11	JQ067650	Arbutus unedo	Ericaceae	145	yes	yes	yes				pseudo	pseudo		no introns		240	intron	
12	NC_034996	Aster altaicus	Asteraceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	234	intron	
13	NC_027434	Aster spathulifolius	Asteraceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	234	intron***	
14	NC_031343	Brasenia schreberi	Cabombaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	234	intron***	
15	NC_016734	Brassica napus	Brassicaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	gone	intron	
16	NC_029889	Carum carvi	Apiaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	234	intron	
17	EF614270	Ceratophyllum demersum	Ceratophyllaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	234	intron	
18	NC_026565	Chloranthus japonicus	Chloranthaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	234	intron	
19	NC_008334	Citrus sinensis	Rutaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	gone	intron	
20	NC_007144	Cucumis sativus	Cucurbitaceae	145	yes	yes	yes				no-5	for		2 introns-rev	yes	153	intron	
21	NC_018541	Elodea canadensis	Hydrocharitaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	240	intron	
22	MF579702	Forsythia suspensa	Oleaceae	145	yes	yes	yes				yes	for		2 introns-rev		234	intron	
23	NC_018109	Gossypium incanum	Malvaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	126	intron	
24	NC_028032	Humulus lupulus	Cannabaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	gone	intron***	
25	NC_034689	Illicium verum	Schisandraceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	234	intron	
26	NC_028617	Juglans regia	Juglandaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	147	intron	
27	DQ400350	Lemna minor	Araceae	145	yes	yes	yes				yes	for		2 introns-rev		gone	intron	
28	NC_026787	Lilium superbum	Liliaceae	145	yes	yes	yes				yes	for		2 introns-rev		pseudo 236	intron	
29	NC_002694	Lotus japonicus	Fabaceae	144/145-p	yes	yes	yes				yes	for		2 introns-rev		gone	intron	
30	NC_020318	Magnolia grandiflora	Magnoliaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	234	intron***	
31	NC_031163	Malus prunifolia	Rosaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	gone	intron	

												accD		clpP1				
	GenBank	Species	Family	Distance	RUA	minus 35	minus 10	rpo genes	ndh genes	accD	accD/CM	direction	clpP1	direction	CT	infA (234)	rpl16 intron	IR perturbations
32	NC_035672	Malus tschonoskii	Rosaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	gone	intron	
33	NC_023256	Melianthus villosus	Melianthaceae	145	yes	yes	yes				yes	for		2 introns-rev		234	intron***	
34	NC_022812	Metapanax delavayi	Araliaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	234	intron	
35	AB237912	Nicotiana sylvestris	Solanaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	318	intron	
36	Z00044	Nicotiana tabacum	Solanaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	318	intron	
37	NC_024542	Nymphaea mexicana	Nymphaeaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	234	intron	
38	GU931818	Olea europaea	Oleaceae	145	yes	yes	yes				yes	for		2 introns-rev		234	intron	
39	NC_034998	Pistacia vera	Anacardiaceae	145	yes	yes	yes				yes	for		2 introns-rev		pseudo 246	intron***	
40	NC_029814	Potamogeton perfoliatus	Potamogetonacea	145	yes	yes	yes				yes	for		2 introns-rev	yes	234	intron	
41	NC_016730	Silene latifolia	Caryophyllaceae	145	yes	yes	yes				no-3(2*)	for		2 introns-rev		gone	intron	regular plastomes
42	NC_016727	Silene vulgaris	Caryophyllaceae	145	yes	yes	yes				no-3(2*)	for		2 introns-rev		gone	intron	regular plastomes
43	NC_035724	Solanum dulcamara	Solanaceae	145	yes	yes	yes				no-5	for		2 introns-rev	yes	169	intron	regular plastomes
44	AJ400848	Spinacia oleracea	Amaranthaceae	145	yes	yes	yes				yes	for		2 introns-rev	yes	177	intron	
45	NC_029813	Tofieldia thibetica	Tofieldiaceae	145	yes	yes	yes				yes	for		2 introns-rev		234	intron	
46	NC_027185	Trillium cuneatum	Melianthaceae	145	yes	yes	yes				yes	for		2 introns-rev		243	intron	
47	NC_029454	Vitis aestivalis	Vitaceae	145	yes	yes	yes				no-2	for		2 introns-rev	yes	234	intron	
48	NC_007957	Vitis vinifera	Vitaceae	145	yes	yes	yes				no-2	for		2 introns-rev	yes	234	intron	
49	NC_015899	Wolffia australiana	Araceae	145	yes	yes	yes				yes	for		2 introns-rev		gone	intron***	
50	NC_015894	Wolffiella lingulata	Araceae	145	yes	yes	yes				yes	for		2 introns-rev		gone	intron***	
		Grasses																
51	NC_031650	Avena sterilis	Poaceae	147	yes	yes	TATACT				gone	n/a		no introns-rev		324*	intron	ycf1/ycf2 genes in IRS gone
52	NC_027476	Hordeum jubatum	Poaceae	147	yes	yes	TATACT				gone	n/a		no introns-rev		342*	intron	ycf1/ycf2 genes in IRS gone
53	NC_022958	Phragmites australis	Poaceae	147	yes	yes	TATACT				gone	n/a		no introns-rev		324*	intron	ycf1/ycf2 genes in IRS gone
54	NC_002762	Triticum aestivum	Poaceae	147	yes	yes	TATACT				gone	n/a		no introns-rev		342*	intron	ycf1/ycf2 genes in IRS gone
55	NC_024175	Oryza glaberrima	Poaceae	147	yes	yes	TATACT				gone	n/a		no introns-rev		324*	intron	ycf1/ycf2 genes in IRS gone

Atypical plastomes

*** gene incorrectly annotated as having the intron missing
Distance = Distance from -10 to 16s RNA
accD/CM = are 5 conserved motifs of Lee et al. 2004 present?
CT = are the 3 amino acid residues of the the catalytic triad present (Wang et al. 1997)
1 IR=only one inverted repeat

	GenBank	Species	Family	Distance	RUA	minus 35	minus 10	rpo genes	ndh genes	accD	accD/CM	accD direction	clpP1	clpP1 direction	CT	infA (234)	rpl16 intron	IR perturbations?	
1	NC_026690	Actinidia chinensis	Actinidiaceae	141	yes	yes	TATCTT				no-5	for	gene gone	n/a		249	intron	2kb IR extension into SSC	
2	NC_035235	Anacardium occidentale	Sapindaceae	152	yes	yes	yes				yes	for		2 introns-rev	yes	120	intron	6.7kb fragment of mito DNA into Ir	
3	NC_030166	Annona cherimola	Annonaceae	162	yes	yes	yes	rpoA highly divergent	short ndhF	696 bp insert	no-5(2)	for	676 and 3rd exor	2 introns-rev	yes	300	intron	IR into SSC and LSC expansions (SS	
4	NC_022457	Berberis bealei	Berberidaceae	145	yes	yes	yes	rpoA lost	ndhK pseudo		yes	for		2 introns-rev	yes***	234	intron***	12 kb expansion into LSC	
5	JQ031013	California macrophylla	Geraniaceae		no	no	no				pseudo	rev	(2 copies w/o intron (1 with 1 intron))			gone	no intron		
6	NC_033337	Campanula punctata	Companulaceae	145	no	yes	yes			gene gone	n/a	n/a	2 copies in IRs	pseudo		pseudo	intron		
7	NC_026203	Campanula takesimana	Companulaceae	145	no	yes	yes			gene gone	n/a	n/a	2 copies in IRs	pseudo		pseudo	intron		
8	NC_027618	Carnegiea gigantea	Cactaceae		no	no	no		pseudo/lost		pseudo	for-reloca	2 copies in LSC	no intron-1 for/1	yes	246	intron***		
9	NC_035236	Cinnamomum verum	Lauraceae	141	yes	yes	yes				yes	for	651	2 introns-rev	yes	234	intron	LSC into Irb	
10	NC_036088	Cyphia belfastica	Companulaceae	145	no	no	yes				pseudo	for		no intron-rev		pseudo	intron		
11	NC_035360	Downingia elegans	Companulaceae	145	no	yes	yes		pseudo/lost	gone	n/a			2 introns-rev		pseudo	intron		
12	NC_029770	Drosera rotundifolia	Droseraceae		no	no	no				no-1(1)-2(1)-4(1)	for		pseudo		420	intron***	Rearranged plastome	
13	NC_015083	Erodium carvifolium	Geraniaceae		no	no	no			gene gone	n/a	n/a		1 intron-rev		gone	no intron	(ycf1, ycf2 lost)-1 IR	
14	NC_027065	Erodium chrysanthum	Geraniaceae		no	no	no		pseudo/lost		no-3(2)-4(3)-5(2)	rev		no intron-for		gone	no intron	(ycf1, ycf2 lost)	
15	NC_025906	Erodium crassifolium	Geraniaceae		no	no	no				no-3(2)-4(2)-5(2)	for		1 intron-rev		pseudo	no intron	(ycf1, ycf2 lost)-1 IR	
16	NC_025907	Erodium gruinum	Geraniaceae		no	no	no		pseudo/lost		no-3(2)-4(2)-5(2)	for		no intron-rev		gone	no intron	(ycf1, ycf2 lost)	
17	NC_014569	Erodium texanum	Geraniaceae		no	no	no			gene gone	n/a	n/a	2 copies in LSC	pseudo-1 intron		gone	no intron	1 IR	
18	NC_025652	Genlisea margaretae	Lentibulariaceae	137	yes	yes	yes	sig. inc. in dn/ds in rpoA	pseudo/lost	gene gone	n/a			2 introns-rev		gone	intron		
19	HM125537	Geranium palmatum	Geraniaceae		no	no	no				no-3(2)-4(2)-5(2)	rev	no intron-for	pseudo		gone	no intron and 2 pseudo co	11 kb (ycf1, ycf2 lost)	
20	NC_024732	Hanabusaya asiatica	Compunulaceae	145	no	yes	yes			gene gone	n/a	n/a	2 copies in IRs	pseudo		gone	intron		
21		Hydrilla verticillata	Hydrocharitaceae		no	no	no				yes	for-at rearrangement breal		2 introns-rev	yes	369-2microsa	intron		
22	DQ673255	Jasminum nudiflorum	Oleaceae		no	no	no			gene gone	n/a	n/a		no intron-rev		270-poly C ru	intron		21 kb inversi
23	NC_034691	Jasminum tortuosum	Oleaceae		no	no	no			gene gone	n/a	n/a		no intron-rev		264-poly C ru	intron		21 kb inversi
24	NC_026703	Ipomoea batatas	Solanaceae	139	yes	yes	yes				yes	for		2 introns-rev***		gone	intron	ycf1 lost	
25	NC_035225	Lychnis wilfordii	Caryophyllaceae		yes	no	no			gene gone	n/a	n/a		2 introns-rev		gone	intron		
26	NC_029830	Monsonia marlothii	Geraniaceae	141	no	yes	no			gene gone	n/a	n/a		no intron-for		gone	no intron	(ycf1, ycf2 lost)-1 IR	
27	HM125538	Monsonia speciosa	Geraniaceae		no	no	no			gene gone	n/a	n/a		no intron-rev		gone	no intron	(ycf1, ycf2 lost)-1 IR	
28		Najas major	Hydrocharitaceae		no	no	no				yes	for		2 introns-rev	no	327	no intron		
29	NC_010361	Oenothera biennis	Onagraceae	74	yes	yes	no				yes	for		no intron-rev		pseudo	intron		
30	NC_010360	Oenothera glazioviana	Onagraceae	74	yes	yes	no				yes	for		no intron-rev		pseudo	intron		
31	NC_010362	Oenothera parviflora	Onagraceae	74	yes	yes	no				yes	for		no intron-rev		pseudo	intron		

												accD		clpP1					
	GenBank	Species	Family	Distance	RUA	minus 35	minus 10	rpo genes	ndh genes	accD	accD/CM	direction	clpP1	direction	CT	infA (234)	rpl16 intron	IR perturbations?	
32	NC_034285	Passiflora edulis	Passifloraceae		no	Yes	Yes				pseudo	rev		no intron-For	yes	gone	intron		
33	NC_023261	Pelargonium alternans	Geraniaceae		no	no	no				no-3(2)-4(2)	for		2 introns-rev		gone	no intron-dup		
34	NC_031194	Pelargonium citronellum	Geraniaceae		no	no	no	rpoA highly divergent			no-3(2)-4(2)	for		2 introns-rev		gone	no intron-dup		
35	NC_031195	Pelargonium dolomiticur	Geraniaceae		no	no	no				pseudo	rev	2 copies in Irs	no intron		gone	no intron-dup		
36	NC_031197	Pelargonium echinatum	Geraniaceae		no	no	no				no-3(2)-4(2)	for		2 introns-rev		gone	no intron-dup		
37	DQ897681	Pelargonium x hortorum	Geraniaceae		no	no	no	3 rpoA orfs		gene gone	n/a	n/a	2 copies in Irs	2 introns		pseudo - F dir	no intron+2 copies in Irs	76 kb	
38	NC_014057	Pisum sativum	Fabaceae		no	no	no				no-5(1)	rev		1 intron-For		gone	intron	1 IR	
39	NC_009143	Populus trichocarpa	Salicaceae	140	yes	yes	yes				no-2(1)	for		2 introns-rev		pseudo	intron		
40	NC_029815	Sagittaria lichuanensis	Alismataceae		no	no	no				yes	for		2 introns-rev	yes	255	intron		
41	NC_023359	Silene chalcedonica	Caryophyllaceae		yes	no	no				no-3(2*)-5(1*)	rev	2 copies in LSC	no intron		gone	intron	inversions and transpositions in th	
42	NC_016729	Silene conica	Caryophyllaceae		no	no	yes				no-3(2*)	rev		no intron		gone	intron	inversions	
43	NC_016728	Silene noctiflora	Caryophyllaceae		no	no	yes				pseudo	rev		no intron		gone	no intron	most complicated plastome in Cary	
44	NC_010442	Trachelium caeruleum	Companulaceae	145	no	yes	yes				2 pseudo copies - in within y	2 copies in IRs		pseudo??		pseudo	intron (operon moved to 19,000)		
45	NC_020372	Trithuria inconspicua	Hydatellaceae	145	no	no	no				no	for		1 intron-rev***	yes	249	intron	IR expansion into SSC	
46	NC_021426	Trochodendron aralioide	Trochodendraceae	145	yes	yes	yes				yes	for		2 introns-rev		234	intron		
47	NC_019616	Vaccinium macrocarpon	Ericaceae	145	no	yes	yes				pseudo	rev	gene gone			pseudo	intron***		
48	NC_026778	Vanilla planifolia	Orchidaceae	162	no	yes	yes		pseudo/lost		yes	for		2 introns-rev		246-2 indels	intron		
49	NC_027155	Vicia sativa	Fabaceae	265	no	yes	no				no-3(2)-4(1)-5(1)	rev		1 intron-For		gone	intron	1 IR	
50	NC_036014	Zostera marina	Zosteraceae	140	yes	yes	yes				no	for		2 introns-rev	yes	234	intron***		
51	NC_029712	Cymbidium lancifolium	Orchidaceae	151	yes	yes	yes		pseudo/lost		yes	for		2 introns-rev		234	intron		
52	NC_035336	Dendrobium chrysanthur	Orchidaceae	147	yes	yes	yes		pseudo/lost		yes	for		2 introns-rev		234	intron		
53	NC_035872	Circaeaster agrestis	Ranunculaceae	147	no	yes	no		pseudo/lost	gene gone	n/a	n/a		2 introns-rev		pseudo	intron	rearranged IR and SSC	

Chapter 4 List of appendices

Appendix A. Chloroplast genes omitted from phylogenetic analysis

Appendix B. Four classes of chloroplast functional genes used in phylogenetic analysis

Appendix C. Maximum likelihood and Bayesian inference parameters and estimates

APPENDIX A

accD: Highly divergent in Najas and Hydrilla, at reversal breakpoint in LSC in Hydrilla (Chapter 3). Missing in Nechamandra, Vallisneria, Limnobium, Hydrocharis, Lagarosiphon, and only a small piece in Halophila, and N. guadalupensis in Alismatales dataset. Butomus and Elodea blast 100% of region with greater than 87% sim to lots of other angiospermas, but Najas only blasts 30% of region with 90% sim.

clpP: Potential pseudogene in Najas. Missing in Halophila, Nechamandra, Vallisneria, and truncated in Enhalus. Original annotation of exon1 for N. guadalupensis, Enhalus and Thalassia is incorrect, and three exons concatenated incorrectly in Alismatales dataset. Exon 1 and 3 in Najas gives no BLAST hit. For exon 2, the highest coverage is 80% with 85% identity. No blast hit for exon 3 for Hydrilla (Elodea exon 3 blasts to range of monocots with >87% similarity). No sequence similarity found in DOGMA.

infa: Potential pseudogene in Najas. Missing in Halophila, Enhalus and Thalassia in Alismatales dataset. Vallisneria has no significant blast hit. N. guadalupensis infA gene was duplicated side by side in the matrix. Poor assembly in Hydrilla (Chapter 3), with paralogues.

ycf1: This gene was edited in Gblocks in the original Alismatales dataset, and the reading frame was not preserved. I am not confident in the alignment of Nechamandra and Vallisneria, and Thalassia has a lot of missing data in the middle of the gene and I am unsure why Gblocks did not get rid of this. Divergent even across Najas, with many repeat regions, therefore questionable alignment with Gblocks-trimmed region of the whole datset.

ycf2: As I experienced difficulty in assembling this gene in Hydrilla (Chapter 3), and central regions of this gene are coded as missing for Enhalus, Hydrocharis, Nechamandra, Limnobium and Hydrocharis in the Alismatales dataset, I have removed this gene from the analysis.

ndhA missing/pseudogene in subfamily Hydrilloideae
ndhB missing/pseudogene in subfamily Hydrilloideae
ndhC missing/pseudogene in subfamily Hydrilloideae
ndhD missing/pseudogene in subfamily Hydrilloideae
ndhE missing/pseudogene in subfamily Hydrilloideae
ndhF missing/pseudogene in subfamily Hydrilloideae
ndhG missing/pseudogene in subfamily Hydrilloideae
ndhH missing/pseudogene in subfamily Hydrilloideae
ndhI missing/pseudogene in subfamily Hydrilloideae
ndhJ missing/pseudogene in subfamily Hydrilloideae
ndhK missing/pseudogene in subfamily Hydrilloideae

matK To prevent potential noise, I removed the matK gene for Nechamandra and Vallisneria from the dataset. They are highly divergent in the Alismatales dataset. Assembly queried for Nechamandra, as the sequence does not match Nechamandra alternifolia on Genbank. Additionally, Vallisneria asiatica has a matK pseudogene sequence in genbank that has a greater sequence similarity to the other Hydrocharit seqences here.

APPENDIX B

Genes of four chloroplast functional classes used in phylogenetic analysis

- "_GB" indicates indel regions of alignment stripped in Gblocks
- "_3" indicates 3' end of gene alignment removed manually

Photosynthesis	Ribosomal proteins	Transcription PEP polymerase	Ribosomal RNA
atpA	rpl2_3	rpoA_GB	rrn4.5
atpB_GB	rpl14	rpoB_GB	rrn5
atpE	rpl16_3	rpoC1_GB	rrn16
atpF	rpl20_GB	rpoC2_GB	rrn23
atpH	rpl22_3	matK_3	
atpI	rpl23		
ccsA_GB	rpl32_3		
cemA	rpl33_3		
petA	rpl36		
petB	rps2_GB		
petD_GB	rps3_GB		
petG	rps4_GB		
petL	rps7		
petN	rps8		
psaA	rps11_GB		
psaB	rps12		
psaC	rps14		
psaI	rps15_GB		
psaJ	rps16_GB		
psbA	rps18_GB		
psbB	rps19		
psbC			
psbD_3			
psbE			
psbF			
psbH			
psbI			
psbJ			
psbK			
psbL			
psbM			
psbN			
psbT_3			
psbZ			
rbcL			
ycf3			
ycf4			

APPENDIX C

ML analyses-partition rates unlinked/1000BS-IQTREE	All Genes	Photosynthesis	PEP polymerase	Ribosomal proteins	Ribosomal RNA
Log-likelihood of the tree:	-170728.94	-78486.49	-43877.38	-33685.31	-12901.69
Unconstrained log-likelihood (without tree):	-173207.07	-67675.99	-40804.56	-34457.22	-18870.36
Number of free parameters (#branches + #model parameters):	91	80	68	78	58
Bayesian information criterion (BIC) score:	342434.54	157776.02	88375.87	68067.12	26309.12
Total tree length (sum of branch lengths):	0.51	0.43	0.77	0.73	0.18
Sum of internal branch lengths:	0.25	0.21	0.36	0.37	0.07
% of tree length	48.30%	49.12%	46.67%	50.37%	41.47%

All genes Partition subsets	No. sequences	No. sites	Parsimony informative	ID Model	Speed	Parameters
Subset 1	26	13231	1834	GTR+I+G4	0.7437	GTR{3.70268,3.91229,0.426463,1.56199,4.50504}+F{0.287406,0.193145,0.298392,0.221057}+I{0.423025}+G4{0.805891}
Subset 2	26	13231	1261	GTR+I+G4	0.5256	GTR{1.51539,4.62783,0.286371,1.98295,2.53918}+F{0.269291,0.215657,0.190719,0.324333}+I{0.464165}+G4{0.559532}
Subset 3	26	13231	4468	GTR+G4	2.0268	GTR{1.29538,3.52458,0.108257,1.12113,3.80369}+F{0.315776,0.136052,0.157353,0.390819}+G4{1.05259}
Subset 4	26	6123	268	GTR+I+G4	0.3602	GTR{0.788388,1.95495,0.404285,0.334178,4.38613}+F{0.263895,0.231761,0.313653,0.190692}+I{0.490886}+G4{0.408374}
Total		**45816**	**7831**			
Photosynthesis						
Subset 1	26	7626	604	GTR+I+G4	0.4641	GTR{3.42903,3.4454,0.349356,1.13848,4.95195}+F{0.259217,0.185008,0.31994,0.235835}+I{0.519836}+G4{0.453436}
Subset 2	26	7626	356	GTR+I+G4	0.2735	GTR{2.26721,6.34723,0.282672,3.73449,3.91595}+F{0.2448,0.231852,0.183377,0.339971}+I{0.613851}+G4{0.330268}
Subset 3	26	7626	2521	GTR+G4	2.2623	GTR{1.39907,4.43683,0.150104,1.22998,4.63272}+F{0.289973,0.144085,0.15968,0.406262}+G4{0.962878}
Total		**22878**	**3481**			
PEP polymerase						
Subset 1	26	5235	1026	GTR+G4	0.7548	GTR{2.49613,3.89871,0.275169,1.92145,3.74963}+F{0.305434,0.19492,0.233364,0.266282}+G4{0.554216}
Subset 2	26	4029	1299	GTR+G4	1.3184	GTR{1.15185,2.93257,0.0748243,1.00818,3.00171}+F{0.322744,0.144094,0.167638,0.365524}+G4{0.995438}
Total		**9264**	**2325**			
Ribosomal proteins						
Subset 1	26	2517	533	GTR+G4	0.8736	GTR{2.61566,3.03056,0.445089,1.5048,2.68013}+F{0.367226,0.201582,0.264107,0.167086}+G4{0.504066}
Subset 2	26	2517	414	GTR+G4	0.6726	GTR{1.51205,4.45204,0.439873,1.75067,2.79232}+F{0.294681,0.205279,0.222183,0.277857}+G4{0.347643}
Subset 3	26	2517	810	GTR+G4	1.4538	GTR{1.1734,2.43312,0.0703805,0.925316,2.66847}+F{0.374558,0.120872,0.148981,0.355589}+G4{0.906624}
Total		**7551**	**1757**			
Ribosomal RNA						
Subset 1	26	6123	268	GTR+I	1.0000	GTR{0.812034,2.08877,0.397845,0.361518,4.59525}+F{0.263895,0.231761,0.313653,0.190692}+I{0.774262}
Overall Total		**45816**	**7831**			

MrBayes Mean of 2 runs	4 subsets All Genes	3 subsets Photosynthesis	2 subsets PEP polymerase	3 subsets Ribosomal proteins	1 subset Ribosomal RNA
LnL	-170717.60	-78522.15	-43908.20	-33721.39	-12885.91
LnPr	122.25	125.04	93.02	100.59	122.23
TL(all)	0.51	0.43	0.76	0.73	0.53
alpha1	1.48	1.65	0.57	0.51	
alpha2	0.76	0.59	1.00	0.36	
alpha3	1.06	0.97		0.91	
alpha4	0.51				
pinvar	0.52	0.70			0.77
m1	0.74	0.46	0.75	0.87	0.55
m2	0.53	0.27	1.32	0.67	
m3	2.03	2.26		1.46	
m4	0.36				

MrBayes parameters for full gene dataset:

```
lset applyto=(1,2,4) nst=6 ngammacat=4 rates=invgamma;
lset applyto=(3) nst=6 ngammacat=4 rates=gamma;
prset applyto=(all) statefreqpr=Dirichlet(1.0,1.0,1.0,1.0) ratepr=variable revmatpr=dirichlet(1,1,1,1,1,1) brlenspr=Unconstrained:GammaDir(1.0,0.100,1.0,1.0) shapepr=exponential(1.0);
unlink shape=(all) statefreq=(all) revmat=(all); [Rem: default is branchlengths linked (brlens=linked)]

    mcmcp ngen= 20000000 relburnin=yes burninfrac=0.25 printfreq=2000  samplefreq=2000 nruns=2 nchains=4 savebrlens=yes filename = mcmc;
    mcmc;
    sump;
    sumt;
END;
```